DENSITY APPROXIMATIONS FOR MULTIVARIATE AFFINE JUMP-DIFFUSION PROCESSES

DAMIR FILIPOVIĆ[1], EBERHARD MAYERHOFER[2], AND PAUL SCHNEIDER[3]

ABSTRACT. We introduce closed-form transition density expansions for multivariate affine jump-diffusion processes. The expansions rely on a general approximation theory which we develop in weighted Hilbert spaces for random variables which possess all polynomial moments. We establish parametric conditions which guarantee existence and differentiability of transition densities of affine models and show how they naturally fit into the approximation framework. Empirical applications in credit risk, likelihood inference, and option pricing highlight the usefulness of our expansions. The approximations are extremely fast to evaluate, and they perform very accurately and numerically stable.

1. INTRODUCTION

Most observed phenomena in financial markets are inherently multivariate: stochastic trends, stochastic volatility, and the leverage effect in equity markets are well-known examples. The theory of affine processes provides multivariate stochastic models with a well established theoretical basis and sufficient degree of tractability to model such empirical attributes. They enjoy much attention and are widely used in practice and academia. Among their best-known proponents are Vasicek's interest rate model (Vasicek, 1977), the square-root model Cox et al. (1985), Heston's model (cf. Heston, 1993), and affine term structure models (Duffie and Kan, 1996; Dai and Singleton, 2000; Collin-Dufresne et al., 2008). Affine models owe their popularity and their name to their key defining property: their characteristic function is of exponential affine form and can be computed by solving a system of generalized Riccati differential equations (cf. Duffie et al. (2003)). This allows for computing transition densities and transition probabilities

[1]École Polytechnique Fédérale de Lausanne and Swiss Finance Institute, Quartier UNIL-Dorigny, Extranef 218, CH - 1015 Lausanne, Switzerland

[2]Vienna Institute of Finance, Heiligenstädter Str. 46-48, 1190 Vienna, Austria

[3]Warwick Business School, University of Warwick, Coventry CV4 7AL, United Kingdom

E-mail addresses: damir.filipovic@epfl.ch, eberhard.mayerhofer@vif.ac.at, paul.schneider@wbs.ac.uk.

by means of Fourier inversion (Duffie et al., 2000). Transition densities constitute the likelihood which is an ingredient for both frequentist and Bayesian econometric methodologies.[1] Also, they appear in the pricing of financial derivatives. However, Fourier inversion is a very delicate task. Complexity and numerical difficulties increase with the dimensionality of the process. Efficient density approximations avoiding the need for Fourier inversion are therefore desirable.

This paper is concerned with directly approximating the transition density without resorting to Fourier inversion techniques. We pursue a polynomial expansion approach, an idea that has been proposed by Schoutens (2000), Aït-Sahalia (2002) and Hurn et al. (2008) among others for univariate diffusion processes. Extensions for multivariate (jump-)diffusions do exist in Aït-Sahalia (2008) and Yu (2007), but they follow a different route by approximating the Kolmogorov forward-, and backward partial differential equations. Our approach exploits a crucial property of affine processes. Under some technical conditions, conditional moments of all orders exist and are explicitly given in terms of derivatives of the affine characteristic exponential function, see Duffie et al. (2000). This ensures that the coefficients of the polynomial expansions can be computed without approximation error.

We present a general theory of density approximations with several traits of the affine model class in mind. The assumptions made for the general theory are then justified by proving existence and differentiability of the true, unknown transition densities of affine models. These theoretical results, contrary to the density approximations themselves, do rely on Fourier theory. Specifically we investigate the asymptotic behavior of the characteristic function with novel ODE techniques.

We improve earlier work, along several lines. Our method (i) is applicable to multivariate models; (ii) works equally well for reducible and irreducible processes in the sense of Aït-Sahalia (2008)[2], in particular stochastic volatility models; (iii) produces density approximations the quality of which is independent of the time interval between observations; (iv) allows for expansions on the "correct" state space. That is, the support of the density approximation agrees with the support of the true, unknown transition density as in Hurn et al. (2008) and Schoutens (2000); (v) produces density approximations that integrate to unity by construction, hence are much more amenable to applications that demand the constant of proportionality than the purely polynomial expansions from Aït-Sahalia (2008).[3] A specialization on affine models is not a severe limitation, since virtually any continuous-time multivariate application is based on affine models.[4] This includes Wishart processes Bru (1991) and even general affine matrix-valued processes (Cuchiero et al., 2010a). This paper therefore provides a unified framework for

---

[1]Various other approaches for parameter estimation for discretely observed Markov processes can be found in the literature (excellent comprehensive surveys are for example in Hurn et al., 2007; Sørensen, 2004; Aït-Sahalia, 2007). The approaches range from likelihood approximation using Bayesian data augmentation (Roberts and Stramer, 2001; Elerian et al., 2001; Eraker, 2001; Jones, 1998), estimating functions (Bibby et al., 2004), up to the efficient method of moment Gallant and Tauchen (2009). Only few of them make use of the properties of affine models, however (e.g. Singleton, 2001; Bates, 2006).

[2]A model is said to be *reducible* Aït-Sahalia (2008) if its diffusion function can be transformed one-to-one into a constant)

[3]The Markov chain Monte Carlo sampling schemes from Stramer et al. (2009) accommodate Bayesian likelihood-based inference using expansions from Aït-Sahalia (2008) even in absence of the normalizing constant, but at a high computational cost.

[4]In discrete-time, Le et al. (2010) show how $\mathbb{Q}$-affine models may be constructed to exhibit non-affine dynamics under $\mathbb{P}$.

econometric inference for financial models, because in applications one typically needs to evaluate, both, the transition densities themselves, as well as integrals of payoff functions against the transition densities for model-based asset pricing. This complements the methods recently developed in Chen and Joslin (2011) and Kristensen and Mele (2011), which are aimed at asset pricing only.[5]

The paper proceeds as follows: Section 2 develops a general theory of orthonormal polynomial density approximations in certain weighted $\mathcal{L}^2$ spaces–under suitable integrability and regularity assumptions. These may be validated by the sufficient criteria presented subsequently in Section 3. The density approximations are then specialized within the context of affine processes: Section 4 reviews the affine transform formula and the polynomial moment formula for affine processes, which in turn allows the aforementioned polynomial approximations. The main theoretical contribution–constituted by fairly general results on existence and differentiability of transition densities of affine processes– is elaborated in Section 4.3. In Section 5 we introduce candidate weight functions and the Gram-Schmidt algorithm to compute orthonormal polynomial bases corresponding to these weights, along with important examples. Section 6 relates existing techniques for density approximations to ours. An empirical study is presented in Section 7: applications in stochastic volatility (Section 7.2), credit risk (Section 7.3), likelihood inference (Section 7.4), and option pricing (Section 7.5), support the tractability and usefulness of the likelihood expansions. Section 8 concludes. The proofs of our main results are given in Appendices A–C.

In the paper we will use the following notational conventions. The nonnegative integers are denoted by $\mathbb{N}_0$. The length of a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ is defined by $|\alpha| = \alpha_1 + \cdots + \alpha_d$, and we write $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$ for any $\xi \in \mathbb{R}^d$. The degree of a polynomial $p(x) = \sum_{|\alpha| \geq 0} p_\alpha x^\alpha$ in $x \in \mathbb{R}^d$ is defined as $\deg p(x) = \max\{|\alpha| \mid p_\alpha \neq 0\}$. For the likelihood ratio functions below we define $0/0 = 0$. The class of p-times continuously differentiable (or continuous, if $p = 0$) functions on $\mathbb{R}^d$ is denoted by C^p.

2. DENSITY APPROXIMATIONS

Let g denote a probability density on $\mathbb{R}^d$ whose polynomial moments

$$\mu_\alpha = \int_{\mathbb{R}^d} \xi^\alpha \, g(\xi) \, d\xi$$

of every order $\alpha \in \mathbb{N}_0^d$ exist and are known in closed form. For example, g may denote the pricing density in a financial market model. Typically, g is not known in explicit form, and needs to be approximated. Let w be an auxiliary probability density function on $\mathbb{R}^d$. The aim is to expand the likelihood ratio g/w in terms of orthonormal polynomials of w in order to get an explicit approximation for the unknown density function g. This can be formalized as follows. Define the weighted Hilbert space $\mathcal{L}_w^2$ as the set of (equivalence classes of) measurable functions f on $\mathbb{R}^d$ with finite $\mathcal{L}_w^2$-norm defined by

$$\|f\|_{\mathcal{L}_w^2}^2 = \int_{\mathbb{R}^d} |f(\xi)|^2 \, w(\xi) \, d\xi < \infty.$$

---

[5]It is of course conceivable to mix the mentioned methods. For example, one could use transition densities developed in this paper, while approximating asset prices using the generalized Fourier transform in Chen and Joslin (2011), whenever the payoff function allows it, or the error expansion method from Kristensen and Mele (2011).

Accordingly, the scalar product on $\mathcal{L}_w^2$ is denoted by

$$\langle f, h \rangle_{\mathcal{L}_w^2} = \int_{\mathbb{R}^d} f(\xi)\, h(\xi)\, w(\xi)\, d\xi.$$

We will now proceed under the following assumptions. Sufficient conditions for the assumptions to hold are provided in Section 3 below.

Assumption 1. There exists an orthonormal basis of polynomials $\{H_\alpha \mid \alpha \in \mathbb{N}_0^d\}$ of $\mathcal{L}_w^2$ with $\deg H_\alpha = |\alpha|$. This implies $H_0 = 1$ in particular.

Assumption 2. The likelihood ratio function g/w lies in $\mathcal{L}_w^2$. This is equivalent to $\int_{\mathbb{R}^d} \frac{g(\xi)^2}{w(\xi)}\, d\xi < \infty$.

Consequently, the coefficients

$$c_\alpha = \left\langle \frac{g}{w}, H_\alpha \right\rangle_{\mathcal{L}_w^2} = \int_{\mathbb{R}^d} H_\alpha(\xi)\, g(\xi)\, d\xi \quad (= 1 \text{ for } \alpha = 0)$$

are well defined and given explicitly[6] in terms of the coefficients of H_α and the polynomial moments μ_α of g. Moreover, according to standard $\mathcal{L}_w^2$-theory, the sequence of pseudo-likelihood ratios[7] $1 + \sum_{|\alpha|=1}^J c_\alpha H_\alpha$ approximates the likelihood ratio g/w in $\mathcal{L}_w^2$ for $J \to \infty$. In fact, defining the pseudo-density functions[8]

$$(2.1) \qquad g^{(J)}(x) = w(x)\left(1 + \sum_{|\alpha|=1}^J c_\alpha H_\alpha(x)\right)$$

the following properties can be established.

Theorem 2.1. *The pseudo-density functions* $g^{(J)}$ *satisfy*

$$(2.2) \qquad \int_{\mathbb{R}^d} g^{(J)}(\xi)\, d\xi = 1$$

$$(2.3) \qquad \lim_{J \to \infty} \frac{g^{(J)}}{w} = \frac{g}{w} \quad \text{in } \mathcal{L}_w^2$$

$$(2.4) \qquad \lim_{J \to \infty} \int_{\mathbb{R}^d} \left| g^{(J)}(\xi) - g(\xi) \right|^2 \frac{d\xi}{w(\xi)} = 0.$$

Property (2.2) proves to be very useful for applications where the constant of proportionality is needed, for example option pricing and the computation of Bayes factors.

Proof. A calculation shows that

$$\int_{\mathbb{R}^d} H_\alpha(\xi)\, w(\xi)\, d\xi = \langle H_\alpha, 1 \rangle_{\mathcal{L}_w^2} = \langle H_\alpha, H_0 \rangle_{\mathcal{L}_w^2} = 0,$$

---

[6]This is an advantage over the method in Aït-Sahalia (2002) which also relies on series expansions, where the coefficients are functions of expectations of nonlinear moments, and therefore have to be approximated in general.
[7]See Footnote 8 below for an explanation of this terminology.
[8]Theorem 2.1 below states that $g^{(J)}$ integrates to one, but $g^{(J)}$ may take negative values. Whence we shall call $g^{(J)}$ a pseudo-density function, and $g^{(J)}/w$ a pseudo-likelihood ratio.

by the orthogonality of H_α and $H_0 = 1$. Hence $\int_{\mathbb{R}^d} g^{(J)}(\xi)\, d\xi = \int_{\mathbb{R}^d} w(\xi)\, d\xi = 1$, which proves (2.2). Properties (2.3) and (2.4) are formal restatements of the discussion preceding the theorem.

□

The idea of expanding the likelihood ratio function g/w in orthonormal polynomials of w is simple and powerful. An overview and discussion of related literature can be found e.g. in Bernard (1995). In particular, for the case where w is the standard Gaussian density, (2.1) is actually the Gram–Charlier expansion of g. But note that Assumption 2 is very restrictive in this case. This is why the Gram–Charlier series diverges in most cases of interest, which is sometimes given as an argument against the use of it. However, the blame is on the choice of the Gaussian as auxiliary density. The efficiency of the approximation (2.3), or equivalently (2.4), lies in the appropriate choice of the auxiliary density function w and the corresponding orthonormal polynomials H_α.

Here is a first result towards a good choice of w. The intuition is to choose w as close as possible to the unknown density function g, in the sense that the pseudo-likelihood ratio g/w is close to one. This should be achieved if many of the coefficients c_α, other than $c_0 = 1$, are equal to zero. This will also improve the numerical efficiency of the approximation as the respective orthonormal polynomials H_α need not be computed. Denote the polynomial moments of w by

$$\lambda_\alpha = \int_{\mathbb{R}^d} \xi^\alpha\, w(\xi)\, d\xi.$$

Lemma 2.2 (Moment Matching Principle). *Suppose for some $n \geq 1$, we have $\mu_\alpha = \lambda_\alpha$ for all $|\alpha| \leq n$. Then $c_\alpha = 0$ for $1 \leq |\alpha| \leq n$.*

Proof. The assumption implies that, for $1 \leq |\alpha| \leq n$,

$$c_\alpha = \int_{\mathbb{R}^d} H_\alpha(\xi)\, g(\xi)\, d\xi = \int_{\mathbb{R}^d} H_\alpha(\xi)\, w(\xi)\, d\xi = \langle H_\alpha, 1 \rangle_{\mathcal{L}^2_w} = \langle H_\alpha, H_0 \rangle_{\mathcal{L}^2_w} = 0,$$

by the orthogonality of H_α and $H_0 = 1$. □

3. Sufficient Conditions for Assumptions 1 and 2

In this section we provide sufficient conditions for Assumptions 1 and 2 to hold. The proofs of the following lemmas are postponed to Appendix A. We first provide sufficient conditions on w that guarantee that Assumption 1 is satisfied.

Lemma 3.1. *Suppose that the density function w has a finite exponential moment*

$$(3.1) \qquad \int_{\mathbb{R}^d} e^{\epsilon_0 \|\xi\|}\, w(\xi)\, d\xi < \infty$$

for some $\epsilon_0 > 0$. Then the set of polynomials is dense in $\mathcal{L}^2_w$. Moreover, Assumption 1 is satisfied.

In applications, the auxiliary density function w on $\mathbb{R}^d$ will often be given as product of marginal densities w_i on $\mathbb{R}$. Hence the following modification of Lemma 3.1 will be useful.

Lemma 3.2. *Let $w_1, \ldots, w_d$ be density functions on $\mathbb{R}$ having finite exponential moments*

$$\int_{\mathbb{R}} e^{\epsilon_i |\xi_i|}\, w_i(\xi_i)\, d\xi_i < \infty$$

for some $\epsilon_i > 0$, $i = 1, \ldots, d$. Then the product density $w(\xi) = w_1(\xi_1) \cdots w_d(\xi_d)$ on $\mathbb{R}^d$ admits a finite exponential moment (3.1) *for $\epsilon_0 = \min_i \epsilon_i$. Moreover, let $\{H_j^i \mid j \in \mathbb{N}_0\}$ denote the corresponding orthonormal basis of polynomials of $\mathcal{L}_{w_i}^2(\mathbb{R})$ by $\deg H_j^i = j$, for $i = 1, \ldots, d$, asserted by Lemma 3.1. Then*

$$H_\alpha(\xi) = H_{\alpha_1}^1(\xi_1) \cdots H_{\alpha_d}^d(\xi_d)$$

defines an orthonormal basis of polynomials of $\mathcal{L}_w^2$ with $\deg H_\alpha = |\alpha|$, and Assumption 1 is satisfied.

Assumption 2 is opposite to Assumption 1 in the sense that there we have to bound the auxiliary density function w from below. The following lemmas provide sufficient conditions for Assumption 2 to hold.

Lemma 3.3. *Assume that g is bounded and has a finite exponential moment*

$$\int_{\mathbb{R}^d} e^{\epsilon_0 \|\xi\|} g(\xi) \, d\xi < \infty$$

for some $\epsilon_0 > 0$. If w decays at most exponentially such that

$$(3.2) \qquad \sup_{x \in \mathbb{R}^d} \frac{e^{-\epsilon_0 \|x\|}}{w(x)} < \infty$$

then Assumption 2 is satisfied.

If the support of w and g is contained in a subset $\mathcal{D}$ of $\mathbb{R}^d$, the situation becomes more difficult as one has to control the rate at which w converges to zero at the boundary of the support set. We provide sufficient conditions for the set $\mathcal{D} = \mathbb{R}_+^m \times \mathbb{R}^n$, starting with the scalar case $\mathcal{D} = \mathbb{R}_+$.

Lemma 3.4. *Let $d = 1$ and $p \in \mathbb{N}$. Assume that g is a bounded density with support in $\mathbb{R}_+$ and has a finite exponential moment*

$$\int_0^\infty e^{\epsilon_0 \xi} g(\xi) \, d\xi < \infty$$

for some $\epsilon_0 > 0$. Assume further that g is of class C^p. If w has support in $\mathbb{R}_+$, and decays at most polynomially at zero and exponentially at infinity such that

$$(3.3) \qquad \sup_{x \in [0,1]} \frac{x^{2p}}{w(x)} < \infty \quad and \quad \sup_{x \geq 1} \frac{e^{-\epsilon_0 x}}{w(x)} < \infty$$

then Assumption 2 is satisfied.

The case where $\mathcal{D} = \mathbb{R}_+^m \times \mathbb{R}^n$ is similar, but requires stronger conditions on g and w. We respect the product structure of the domain by writing $g = g(x, y)$ for $x \in \mathbb{R}_+^m$ and $y \in \mathbb{R}^n$. The following tubular neighborhood of the boundary of $\mathcal{D}$

$$\mathcal{I} = \mathbb{R}_+^m \setminus (1, \infty)^m = \{x \in \mathbb{R}_+^m \mid \min_i x_i \leq 1\}$$

is the convenient multivariate generalization of the unit interval from the above scalar case.

Lemma 3.5. *Let $d = m + n$ and $p \in \mathbb{N}$. Assume that $g(x, y)$ is a bounded density with support in $\mathbb{R}_+^m \times \mathbb{R}^n$ and has a finite exponential moment*

$$\int_{\mathbb{R}_+^m} \int_{\mathbb{R}^n} \mathrm{e}^{\,\epsilon_1 \|\xi\| + \epsilon_2 \|\eta\|} \, g(\xi, \eta) \, d\xi \, d\eta < \infty$$

for some ϵ_1, $\epsilon_2 > 0$. Assume further that $g(x, y)$ is of class C^p in x and the p-th partial derivative $\partial_{x_i}^p g(x, y)$ is bounded on $\mathcal{I} \times \mathbb{R}^n$, for all $i = 1, \ldots, m$. If w has support in $\mathbb{R}_+^m \times \mathbb{R}^n$, and decays at most polynomially around the boundary and exponentially at infinity such that

$$(3.4) \qquad \sup_{(x,y) \in \mathcal{I} \times \mathbb{R}^n} \frac{\min_i x_i^p \, \mathrm{e}^{-\epsilon_2 \|y\|}}{w(x, y)} < \infty \quad and \quad \sup_{(x,y) \in (1, \infty)^m \times \mathbb{R}^n} \frac{\mathrm{e}^{-\epsilon_1 \|x\| - \epsilon_2 \|y\|}}{w(x, y)} < \infty$$

then Assumption 2 is satisfied.

We note that the conditions in Lemmas 3.3, 3.4 and 3.5 can be explicitly verified for transition densities of affine processes, see Corollary 4.4 below.

4. Affine Models

The main application of the polynomial density approximation is for affine factor models. In this section, we follow the setup of Duffie et al. (2003), which we now briefly recap. Let $d = m + n \geq 1$. We define the index set $J = \{m + 1, \ldots, d\}$, and write $v_J = (v_{m+1}, \ldots, v_d)$ and $m_{JJ} = (m_{kl})_{k, l \in J}$, for any vector v and matrix m. We consider an affine process X on the canonical state space $\mathcal{D} = \mathbb{R}_+^m \times \mathbb{R}^n$ with generator

$$\begin{aligned}
\mathcal{A}f(x) = \sum_{k,l=1}^d & \left(\mathrm{diag}\,(0, a) + \sum_{i=1}^m x_i \alpha_i \right)_{kl} \frac{\partial^2 f(x)}{\partial x_k \partial x_l} + (b + \beta\,x)^\top \nabla f(x) \\
& + \int_{\mathcal{D}} \left(f(x + \xi) - f(x) - \chi_J(\xi)^\top \nabla_J f(x) \right) m(d\xi) \\
& + \int_{\mathcal{D}} (f(x + \xi) - f(x)) \left(\sum_{i=1}^m x_i \mu_i(d\xi) \right)
\end{aligned}$$

for some appropriate positive semidefinite $n \times n$- and $d \times d$-matrices a and α_i, respectively. Here, with $\mathrm{diag}\,(0, a)$ we denote the block-diagonal $d \times d$-matrix with blocks given by the $m \times m$-zero matrix and a. Moreover, $\chi_J(\xi)$ denotes an $\mathbb{R}^n$-valued continuous and bounded truncation function with $\chi_J(\xi) = \xi_J$ in a neighborhood of the origin $\xi = 0$. For detailed parametric restrictions on $(a, \alpha_i, b, \beta, m, \mu_i)$ we refer the reader to Duffie et al. (2003, Definition 2.6). We assume for simplicity[9] that the jump measures μ_i are of finite variation type with integrable large jumps

$$\int_{\mathcal{D}} \|\xi\| \, \mu_i(d\xi) < \infty, \quad i = 1, \ldots, m.$$

---

[9] At the cost of more technical analysis, the following results could also be proved for the general case of infinite variation jumps μ_i with infinite tail mean.

4.1. **Affine Transform Formula.** The analytical tractability of affine models stems from the fact that the characteristic function of $X_t | X_0 = x$ is explicitly given by the affine transform formula

$$(4.1) \qquad \mathbb{E}\left[e^{\mathrm{i} u^\top X_t} \mid X_0 = x \right] = e^{\phi(t, \mathrm{i} u) + \psi(t, \mathrm{i} u)^\top x}, \quad u \in \mathbb{R}^d, \quad x \in \mathcal{D}$$

where the $\mathbb{C}_-$- and $\mathbb{C}_-^m \times \mathrm{i}\mathbb{R}^n$-valued functions $\phi = \phi(t, \mathrm{i} u)$ and $\psi = \psi(t, \mathrm{i} u)$ solve the generalized Riccati equations, for $i = 1, \ldots, m$,

$$\partial_t \phi = \psi_J^\top a\, \psi_J + b^\top \psi + \int_{\mathcal{D}} \left(e^{\psi^\top \xi} - 1 - \psi_J^\top \chi_J(\xi) \right) m(d\xi),$$

$$\phi(0) = 0,$$

$$(4.2) \qquad \partial_t \psi_i = \psi^\top \alpha_i\, \psi + \mathcal{B}_i\, \psi + \int_{\mathcal{D}} \left(e^{\psi^\top \xi} - 1 \right) \mu_i(d\xi),$$

$$\psi_i(0) = \mathrm{i} u_i,$$

$$\partial_t \psi_J = \mathcal{B}_{JJ} \psi_J,$$

$$\psi_J(0) = \mathrm{i} u_J,$$

where we define $\mathcal{B} = \beta^\top$ and write $\mathcal{B}_i$ for the ith row vector of $\mathcal{B}$. Obviously, we have

$$\psi_J(t, \mathrm{i} u) = \mathrm{i} e^{\mathcal{B}_{JJ} t}\, u_J,$$

and $\phi(t, \mathrm{i} u)$ is given by simple integration of the right hand side of its equation.

4.2. **Polynomial Moments.** It is well known that if $X_t | X_0 = x$ has finite k-th moment,

$$\mathbb{E}\left[\|X_t\|^k \mid X_0 = x \right] < \infty \quad \text{for all } x \in \mathcal{D},$$

then $\phi(t, u)$ and $\psi(t, u)$ are of class C^k in u. Moreover, the polynomial moments are explicitly given in terms of the respective mixed derivatives of the characteristic function

$$\mathbb{E}\left[X_t^\alpha \mid X_0 = x \right] = -\mathrm{i}^{|\alpha|} \frac{\partial^{|\alpha|}}{\partial u_{\alpha_1} \cdots \partial u_{\alpha_d}} e^{\phi(t, \mathrm{i} u) + \psi(t, \mathrm{i} u)^\top x}\big|_{u=0}$$

for $|\alpha| \le k$, see e.g. Duffie et al. (2003, Lemma A.1). It follows by inspection that the right hand side of this equation is a real polynomial in x of degree less than or equal to $|\alpha|$. Recently, generalizing the recursive method used in Forman and Sørensen (2008) for Pearson-type diffusions, Cuchiero et al. (2010b) proposed an alternative method to compute the coefficients of this polynomial. The idea rests on the insight that the affine generator $\mathcal{A}$ formally maps $\mathcal{P}_k$ into $\mathcal{P}_k$, where $\mathcal{P}_k$ denotes the finite-dimensional linear space of all polynomials in $x \in \mathbb{R}^d$ of degree less than or equal to k.[10] The generator $\mathcal{A}$ thus restricts to a linear operator $\mathcal{A}_k$ on $\mathcal{P}_k$. Consequently, we obtain the formal representation

$$\mathbb{E}\left[X_t^\alpha \mid X_0 = x \right] = e^{\mathcal{A}_k t}\, x^\alpha$$

where $e^{\mathcal{A}_k t} = \sum_{j \ge 0} \frac{(\mathcal{A}_k t)^j}{j!}$ is the exponential of $\mathcal{A}_k t$. This can be expressed as a matrix. We shall illustrate this for $d = 1$. The dimension of $\mathcal{P}_k$ then equals $k + 1$, and we can pick as canonical

---

[10]This method is not restricted to affine processes, but can be defined for any Markov process with finite k-th moments, and whose infinitesimal generator maps $\mathcal{P}_k$ into itself.

basis of $\mathcal{P}_k$ the set $\mathcal{Q} = \{1, x, \ldots, x^k\}$. For every $j = 0, \ldots, k$ we then calculate symbolically the coefficients q_{ij} in

$$(4.3) \qquad \mathcal{A}_k x^j = \mathcal{A} x^j = \sum_{i=0}^{k} q_{ij} x^i.$$

Hence $\mathcal{A}_k$ can be represented by the upper-triangular matrix $Q = (q_{ij})$ with respect to the basis $\mathcal{Q}$. In other words, if we identify a generic polynomial $p(x) = \sum_{i=0}^{k} p_i x^i$ in $\mathcal{P}_k$ with the vector of its coefficients $p = (p_0, \ldots, p_k)^\top$, then $\mathcal{A}_k p(x) \in \mathcal{P}_k$ equals the polynomial with coefficient vector Qp. Moreover,

$$(4.4) \qquad \mathbb{E}\left[p(X_t) \mid X_0 = x\right] = e^{Qt} p.$$

See Examples 7.1 and 7.2 below for some concrete applications for $d = 1$ and $d = 2$.

4.3. **Existence and Properties of Affine Transition Densities.** In this section we present our main theoretical results, which establish existence and smoothness of the density of the conditional distribution of the affine process X. Moreover, we provide explicitly verifiable conditions asserting that Lemmas 3.3–3.5 apply.

Our first result provides sufficient conditions for the existence and smoothness of a density of $X_t|X_0 = x$, for some $t > 0$ and $x \in \mathbb{R}_+^m \times \mathbb{R}^n$. These easy to check conditions apply in particular to multi-factor affine term-structure models on $\mathbb{R}_+^m \times \mathbb{R}^n$ from Duffie and Kan (1996) and Dai and Singleton (2000), and Heston's stochastic volatility model. The proof is given in Appendix B.

Theorem 4.1. *Assume that the $d \times (n+1)d$-matrix*[11]

$$(4.5) \qquad \mathcal{K} = \left[\sum_{i=1}^{m} \alpha_i, \, \mathrm{diag}\,(0, a), \, \mathrm{diag}\left(0, \mathcal{B}_{JJ}^\top a\right), \ldots, \mathrm{diag}\left(0, (\mathcal{B}_{JJ}^{n-1})^\top a\right) \right]$$

has full rank. Further, let p be a nonnegative integer with

$$(4.6) \qquad p < \min_{i \in \{1, \ldots, m\}} \frac{b_i}{\alpha_{i,ii}} - 1.$$

Then $X_t|X_0 = x$ admits a density $g(\xi)$ of class C^p with support in $\mathbb{R}_+^m \times \mathbb{R}^n$ and the partial derivatives of $g(\xi)$ of orders $0, \ldots, p$ tend to 0 as $\|\xi\| \to \infty$.

We note that condition (4.6) is sharp and cannot be relaxed in general. Consider for instance the scalar square-root diffusion X on $\mathbb{R}_+$ with generator $\mathcal{A}f(x) = \alpha x f''(x) + b f'(x)$. It is well known that for any parameter values $\alpha > 0$ and $b \geq 0$, the distribution of $\frac{2X_t}{\alpha t} \mid X_0 = x$ is noncentral χ^2 with $\frac{2b}{\alpha}$ degrees of freedom and noncentrality parameter $\frac{2x}{\alpha t}$, see e.g. Filipović (2009, Exercise 10.9). The corresponding density function $g(\xi)$ satisfies $\lim_{\xi \to 0} g(\xi) = 0$, and is therefore of class C^0, if and only if the degrees of freedom $\frac{2b}{\alpha} > 2$, see Johnson et al. (1995, Chap. 29). This is exactly what condition (4.6) states for $p = 0$.

As regards exponential moments of $X_t|X_0 = x$, we combine and rephrase some results from Duffie et al. (2003):

---

[11]Here, for given $d \times d$-matrices $B_1, B_2, \ldots, B_n$ the expression $[B_1, B_2, \ldots, B_n]$ denotes the $d \times nd$-block matrix we obtain by putting the matrices next to each other.

Theorem 4.2. *Assume that the jump measures admit exponential moments*

$$\int_{\{\|\xi\|>1\}} e^{q^\top \xi} \, m(d\xi) < \infty \quad and \quad \int_{\{\|\xi\|>1\}} e^{q^\top \xi} \, \mu_i(d\xi) < \infty, \quad i = 1, \dots, m$$

for all q in some open neighborhood V of 0 in $\mathbb{R}^d$. Then the right hand side of (4.2) is analytic in $\psi \in V$. Suppose further that (4.2) admits a V-valued solution $\psi(t, u)$ with $\psi(0, u) = u$ for all $t \in [0, T)$ and for all u in $[-\epsilon_1, \epsilon_1]^m \times [-\epsilon_2, \epsilon_2]^n$, for some $\epsilon_1, \epsilon_2 > 0$. Then $X_t | X_0 = x$ has a finite exponential moment

$$(4.7) \qquad\qquad \mathbb{E}\left[e^{\epsilon_1 \|Y_t\| + \epsilon_2 \|Z_t\|} \mid X_0 = x \right] < \infty$$

for all $t \in [0, T)$, where we denote $Y_t = (X_{1,t}, \dots, X_{m,t})^\top$ and $Z_t = (X_{m+1,t}, \dots, X_{d,t})^\top$.

Proof. That the right hand side of (4.2) is analytic in $\psi \in V$ follows from Duffie et al. (2003, Lemma 5.3). From Duffie et al. (2003, Theorem 2.16 and Lemma 6.5) we then infer that

$$\mathbb{E}\left[e^{q^\top X_t} \mid X_0 = x \right] < \infty$$

for all $t \in [0, T)$ and for all $q \in [-\epsilon_1, \epsilon_1]^m \times [-\epsilon_2, \epsilon_2]^n$. Combining this with the elementary inequality

$$e^{\epsilon_1 \|Y_t\| + \epsilon_2 \|Z_t\|} \le e^{\epsilon_1 \sum_{i=1}^m |X_{it}| + \epsilon_2 \sum_{i=m+1}^d |X_{it}|} \le \sum_{|\alpha|=0}^1 e^{q_\alpha^\top X_t},$$

where we denote $q_\alpha = ((-1)^{\alpha_1} \epsilon_1, \dots, (-1)^{\alpha_m} \epsilon_1, (-1)^{\alpha_{m+1}} \epsilon_2, \dots, (-1)^{\alpha_d} \epsilon_2)^\top \in [-\epsilon_1, \epsilon_1]^m \times [-\epsilon_2, \epsilon_2]^n$, proves (4.7). $\qquad\square$

Note that if $m(d\xi)$ and $\mu_i(d\xi)$ have light tails of the order $e^{-r\|\xi\|^2} d\xi$ for some $r > 0$, or have compact support in particular, then the first assumption of Theorem 4.2 is satisfied for $V = \mathbb{R}^d$. Even then, however, the solution $\psi(t, u)$ exists only on a finite time horizon $t < T < \infty$ for any nonzero $u \in \mathbb{R}^d$ in general. We refer to the discussion of the diffusion case in Filipović and Mayerhofer (2009), see also Filipović (2009, Chapter 10).

We further present an additional result which concerns the existence of the marginal transition density of integrated affine jump-diffusions, which are not covered by Theorem 4.1. In fact, if X is a one-dimensional affine process on $\mathbb{R}_+$, then the two-dimensional process $(dX, X\, dt)^\top$ is affine again, with state space $\mathbb{R}_+^2$. However, its diffusion matrix is degenerate and thus violates the conditions of Theorem 4.1. Nevertheless, a slight adaption of its proof yields the existence of the marginal transition density of the integrated process $\int X\, dt$ under some more stringent conditions. The proof of the following theorem is given in Appendix C.

Theorem 4.3. *Let X be an $\mathbb{R}_+$-valued affine process with parameters $(a = 0, \alpha, b, \beta, m, \mu)$. Further, let p be a nonnegative integer with*

$$(4.8) \qquad\qquad p < \frac{b}{2\alpha} - 1.$$

Then $\int_0^t X_s ds \mid X_0 = x$ admits a a density $g(\xi)$ of class C^p with support in $\mathbb{R}_+$ and the partial derivatives of $g(\xi)$ of orders $0, \dots, p$ tend to 0 as $\xi \to \infty$.

From the application point of view, we can rephrase the statements of the preceding theorems as follows:

Corollary 4.4. *Theorems 4.1–4.3 provide conditions in terms of the parameters of the affine process X such that the assumptions in Lemmas 3.3–3.5, and thus eventually the validity of Assumptions 1 and 2, can explicitly be verified for the density of the (marginal) transition distributions of X.*

5. Examples of Auxiliary Density Functions

For our applications of the polynomial density approximations to affine models we shall make the following specific choices for the auxiliary density function w. For positive coordinates we use the Gamma density

$$(5.1) \qquad \gamma(\xi; D) = \frac{e^{-\xi}\xi^D}{\Gamma[1+D]}$$

of a $\Gamma(1+D, 1)$-distributed random variable. Here, $\Gamma[\cdot]$ denotes the Gamma function. It is easily seen that conditions (3.1)–(3.4) are satisfied for the appropriate parameters p, ϵ_0 and ϵ_1, respectively.

For real-valued coordinates we employ the bilateral Gamma density from Küchler and Tappe (2008a). The corresponding family of distributions nests, for example, the Variance Gamma distribution as a special case. It has very flexible shapes (Küchler and Tappe, 2008b). For the purposes of this paper we make use of a constrained, standardized version with mean centered at zero, unit variance, zero skewness, and excess kurtosis $C > 0$. We denote this standardized bilateral Gamma distribution with $\Gamma_b(C)$. Its characteristic function is given by

$$\Phi_{\Gamma_b}(u; C) = 216^{\frac{1}{C}} \left(\frac{1}{6-Cu^2}\right)^{3/C}, \quad u \in \mathbb{R}, \quad C \in \mathbb{R}_+.$$

By Küchler and Tappe (2008b, eq. 3.6) the bilateral Gamma distribution has a density given by

$$(5.2) \qquad \gamma_b(\xi; C) = \frac{2^{\frac{3(C-2)}{4C}} 3^{\frac{C+6}{4C}} C^{-\frac{C+6}{4C}} |\xi|^{\frac{3}{C}-\frac{1}{2}} K_{\frac{3}{C}-\frac{1}{2}}\left(\frac{\sqrt{6}|\xi|}{\sqrt{C}}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{3}{C}\right)},$$

where $K_n(\xi)$ denotes the modified Bessel function of the second kind. It follows from Küchler and Tappe (2008b, Section 6) that conditions (3.1)–(3.4) are satisfied for the appropriate parameters ϵ_0 and ϵ_2, respectively. The special case with excess kurtosis $C = 1/3$ leads to the following simple expression for the density $\gamma_b(\xi) = \gamma_b(\xi\,; C = 1/3)$, since for half-integer indices the modified Bessel functions evaluate to elementary functions

$$(5.3) \quad \gamma_b(\xi) = \frac{27 e^{-3\sqrt{2}|\xi|}}{1146880\sqrt{2}} \cdot \Big(7776\sqrt{2}|\xi|^7 + 83160\sqrt{2}|\xi|^5 + 180180\sqrt{2}|\xi|^3$$

$$+ 75075\sqrt{2}|\xi| + 1296\xi^8 + 45360\xi^6 + 207900\xi^4 + 210210\xi^2 + 25025\Big).$$

Orthonormal polynomial bases can be constructed for any auxiliary density function w which has finite exponential moment (3.1) by the following Gram-Schmidt process, which is also used in the proof of Lemma 3.1.

Algorithm 5.1 (Gram-Schmidt Process).

$$H_0 = 1$$

$$\widetilde{H}_\alpha = \xi^\alpha - \sum_{0 \le |\beta| \le |\alpha|, \, \beta \ne \alpha} \langle \xi^\alpha, H_\beta(\xi) \rangle_{\mathcal{L}^2_w} H_\beta(\xi)$$

$$H_\alpha = \frac{\widetilde{H}_\alpha}{\left\| \widetilde{H}_\alpha(\xi) \right\|_{\mathcal{L}^2_w}} \qquad \text{(normalization)} \; .$$

Notice that $\deg H_\alpha = \deg \xi^\alpha = |\alpha|$, which is due to the linear independence of the set of monomials $\{ \xi^\alpha \mid \alpha \in \mathbb{N}_0^d \}$. Below are the first five orthonormal polynomials for the Gamma and the bilateral Gamma densities γ and γ_b introduced above.

Example 5.1. The non-normalized orthogonal polynomials for the Gamma density γ are the generalized Laguerre polynomials, the first five of which are

(5.4)
$$\begin{aligned}
&H_0^\gamma(\xi) = 1, \\
&\widetilde{H}_1^\gamma(\xi) = -\xi + D + 1 \\
&\widetilde{H}_2^\gamma(\xi) = \frac{1}{2} \left(\xi^2 - 2\xi(D+2) + D^2 + 3D + 2 \right) \\
&\widetilde{H}_3^\gamma(\xi) = \frac{1}{6} \left(-\xi^3 + 3\xi^2(D+3) - 3\xi \left(D^2 + 5D + 6 \right) + D^3 + 6D^2 + 11D + 6 \right) \\
&\widetilde{H}_4^\gamma(\xi) = \frac{1}{24} (\xi^4 - 4(D+4)\xi^3 + 6(D+3)(D+4)\xi^2 \\
&\qquad\qquad - 4(D+2)(D+3)(D+4)\xi + (D+1)(D+2)(D+3)(D+4))n.
\end{aligned}$$

The normalization constants are given by

(5.5)
$$HO_n^\gamma = \left\| \widetilde{H}_n^\gamma(\xi) \right\|_{\mathcal{L}^2_\gamma} = \sqrt{\frac{\prod_{i=1}^n (i+D)}{n!}}.$$

Example 5.2. For the standardized bilateral Gamma density γ_b in (5.2), the first five non-normalized orthogonal polynomials are

(5.6)
$$\begin{aligned}
&H_0^{\gamma_b}(\xi) = 1 \\
&\widetilde{H}_1^{\gamma_b}(\xi) = \xi \\
&\widetilde{H}_2^{\gamma_b}(\xi) = \xi^2 - 1 \\
&\widetilde{H}_3^{\gamma_b}(\xi) = (-C - 3)\xi + \xi^3 \\
&\widetilde{H}_4^{\gamma_b}(\xi) = -\frac{2 \left(5C^2 + 21C + 18 \right) \left(\xi^2 - 1 \right)}{3(C+2)} - C + \xi^4 - 3,
\end{aligned}$$

and the corresponding normalization constants $HO_n^{\gamma_b} = \left\| \widetilde{H}_n^{\gamma_b}(\xi) \right\|_{\mathcal{L}^2_{\gamma_b}}$ are given by

(5.7)
$$
\begin{aligned}
HO_0^{\gamma_b} &= 1 \\
HO_1^{\gamma_b} &= 1 \\
HO_2^{\gamma_b} &= \sqrt{C+2} \\
HO_3^{\gamma_b} &= \sqrt{\frac{7C^2}{3} + 9C + 6} \\
HO_4^{\gamma_b} &= \sqrt{\frac{2\left(55C^4 + 363C^3 + 822C^2 + 756C + 216\right)}{9(C+2)}}.
\end{aligned}
$$

Example 5.3 (Product Measure). Define the product density $w_{\gamma\gamma_b}$ with support on $\mathbb{R}_+ \times \mathbb{R}$ by

$$ w_{\gamma\gamma_b}(\xi_1, \xi_2; C, D) = \gamma(\xi_1, D)\gamma_b(\xi_2, C) $$

with Gamma density γ defined in (5.1) and bilateral Gamma density γ_b defined in (5.2). Combining Lemma 3.2 and Examples 5.1 and 5.2, we obtain the corresponding orthonormal basis of polynomials $\left\{ H_{n_1}^{\gamma} \cdot H_{n_2}^{\gamma_b} \mid (n_1, n_2) \in \mathbb{N}_0^2 \right\}$.

6. Relation to Existing Approximations

In this section we recall facts about closed-form density approximations from previous literature and relate them to the density expansions of the present paper. A short summary of the capabilities and limitations of the different methods is reported in Table 1.

The closest methodology to the one introduced in Section 2 Aït-Sahalia (2002). One of the key steps is to transform the original process in such a way that a Gaussian-weighted $\mathcal{L}^2$ expansion converges. This is motivated by an analogy to the central limit theorem (see also Aït-Sahalia and Yu, 2005, Introduction): the sampling interval (time between observations) Δ plays the role of the sample size n in the central theorem; conditional on a correct standardization a $N(0,1)$ density turns out to be the correct limiting distribution as $n \to \infty$ (in the central limit theorem) and as $\Delta \to 0$ (for the stochastic process). The correcting Hermite polynomials (the pseudo likelihood ratio) then account for the fact that Δ is not 0. Aït-Sahalia (2002) applies two transformations. The first change of variables yields a unit diffusion process through the Lamperti transform. For univariate diffusions it can be shown that such a transformation always exists. This step introduces nonlinearities in the drift. The resulting process is then centered, and scaled in time. Consequently, a Gaussian-weighted $\mathcal{L}^2$ expansion converges *uniformly* to the true, unknown transition density. This strong convergence result–which of course exceeds the mere $\mathcal{L}^2$ convergence– is proved by using a representation of the true, unknown, transition density in terms of a Brownian bridge functional from Rogers (1985). Due to the nonlinearities in the drift, the coefficients of the Aït-Sahalia (2002) Hermite expansion are generally not known in closed form, however. In practice they are approximated using a Taylor expansion in time in terms of the infinitesimal generator of the process. This is a key difference to the setting of the present paper, where expansions are constructed precisely such that their coefficients are linear in polynomial moments–and those are known for affine processes without approximation error.

In the multivariate case, however, a Lamperti transform is rarely possible, since most applications call for stochastic volatility models which are irreducible (Aït-Sahalia, 2008, Proposition

	Approximations				
	AS02	ASY05	Y07	AS08	this paper
multivariate	No	Yes	Yes	Yes	Yes
everywhere positive	No	No	No	Yes	No
integrates to one	No	No	No	No	Yes
jumps	No	Yes	Yes	No	Yes

TABLE 1. **Comparison of Closed-Form Transition Density Approximations.** AS02 refers to Aït-Sahalia (2002), ASY05 to Aït-Sahalia and Yu (2005), Y07 to Yu (2007), and AS08 to Aït-Sahalia (2008).

1). An entirely different strategy is therefore pursued in Aït-Sahalia (2008) for the irreducible multivariate case, where the log likelihood is expanded in, both, time, and space, so that the coefficients of the expansion may be computed from the Kolmogorov forward and backward equations. This approach is adopted by Yu (2007) with the difference that he also considers jump-diffusions and approximates the transition density itself, rather than the log transition density.

The saddlepoint approach in Aït-Sahalia and Yu (2005) is fundamentally different. It (approximately) solves the Fourier inversion problem by expanding the cumulant generating function about the saddlepoint[12], rather than making use of the Kolmogorov forward and backward equations. The maintained assumption here is that the cumulant generating function is available, even though for diffusions Aït-Sahalia and Yu (2005, Section 4) circumvent this problem by using a Taylor series expansion for small times along the lines of Aït-Sahalia (2002) for nonlinear moments. Though the saddlepoint approach and this paper both facilitate expansion techniques, the objects of the expansion are different and the formulae are unrelated. Saddlepoint approximations are extremely accurate even for low orders (Aït-Sahalia and Yu, 2005, Fig. 2). The price to be paid for this precision is the computational burden of having to solve numerically for the saddlepoint for every pair of forward and backward variables.

7. Applications

In the following we present applications which highlight the usefulness of the transition density approximations developed in this paper. For the empirical investigations considered below we find that there is a trade-off between numerical accuracy and the order of the expansion. Higher-order expansions may perform worse than low-order expansions due to numerical errors that are induced by the limited numeric precision of the computer environment in representing very large or very small numbers. As a general guideline we suggest matching as many moments (cumulants) as possible when choosing $\mathcal{L}^2$ weights, and stopping the expansion at a relatively low order such as $J = 4$. For the present section we adopt notation used conventionally in finance and econometrics. In particular we deviate from Duffie et al. (2003) notation. The time interval between between observations is generally denoted by Δ.

---

[12]For a stochastic process X denote by $K(t, u \mid x_0) = \log \mathbb{E}\left[e^{u^\top X_t} \mid X_0 = x_0\right]$ the cumulant generating function. Suppose $X_t|X_0 = x_0$ has an absolutely continuous law. For any state x the saddlepoint is defined as the solution $\hat{u} = \hat{u}(t, x, x_0)$ in u to the implicit equation $\partial_u K(t, u \mid x_0) = x$.

We strongly recommend checking the above theoretical foundations for the validation of Assumptions 1 and 2 in numerical applications, as outlined in Example 7.1 below.

7.1. Basic Affine Jump-Diffusion (BAJD).

We first consider a square-root process with exponentially distributed jumps. This process has been recently used in papers that study portfolio credit risk (Duffie and Garleanu, 2001; Mortensen, 2006; Eckner, 2009; Feldhütter, 2008) where it is termed basic affine jump-diffusion (BAJD), and in a bivariate form in single-name credit (Schneider et al., 2010). It can be described in SDE form

$$(7.1) \qquad dY_t = (\kappa\theta - \kappa Y_t)\, dt + \sigma\sqrt{Y_t} dW_t + dK_t.$$

The intensity of the compound Poisson process K is $l \geq 0$, and the expected jump size of the exponentially distributed jumps is $\nu \geq 0$. The set of parameters we denote by $\varrho_Y = \{\kappa\theta, \kappa, \sigma, \nu, l\}$ and the domain of the process $\mathcal{D} = \mathbb{R}_+$. By Theorem 4.1, $2\kappa\theta > \sigma^2$ ensures existence of transition densities.

Example 7.1 (Developing an $\mathcal{L}^2_\gamma$ Expansion for the BAJD)**.** To exemplify the necessary steps to develop a density expansion, we consider here an explicit example and compute an order $J = 4$ density expansion for the BAJD process from eq. (7.1). The $\mathcal{L}^2$ weight we use here is a Gamma distribution $\Gamma(1 + D, 1)$ with density function γ from eq. (5.1).
Step 1. Computing Conditional Moments: The generator of the BAJD is

$$\mathcal{A}f(x) = (\kappa\theta - \kappa x)\frac{\partial f(x)}{\partial x} + \frac{1}{2}\sigma^2 x\frac{\partial^2 f(x)}{\partial x^2} + l\int_{\mathbb{R}_+}(f(x + \xi) - f(x))\frac{1}{\nu}e^{-\frac{\xi}{\nu}}d\xi.$$

Hence the matrix $Q = (q_{ij})$ from (4.3) relative to the canonical basis $\{1, x, x^2, x^3, x^4\}$ equals

$$Q = \begin{pmatrix} 0 & \kappa\theta + l\nu & 2l\nu^2 & 6l\nu^3 & 24l\nu^4 \\ 0 & -\kappa & \sigma^2 + 2\kappa\theta + 2l\nu & 6l\nu^2 & 24l\nu^3 \\ 0 & 0 & -2\kappa & 3\sigma^2 + 3\kappa\theta + 3l\nu & 12l\nu^2 \\ 0 & 0 & 0 & -3\kappa & 6\sigma^2 + 4\kappa\theta + 4l\nu \\ 0 & 0 & 0 & 0 & -4\kappa \end{pmatrix}.$$

Note the upper-triangular form. A symbolic mathematics software package such as Mathematica or Maple will be able to compute the matrix exponential $e^{Q\Delta}$ in closed form. The conditional moments $\mu_n(y_0, \Delta, \varrho_Y) = \mathbb{E}\left[Y_\Delta^n \mid Y_0 = y_0, \varrho_Y\right]$ may then be obtained by plugging into formula (4.4). We obtain the conditional moments as polynomials of order ≤ 4 in the backward variable y_0. Below we will suppress dependence of the moments on y_0, Δ, ϱ_Y to lighten notation.
Step 2. Scaling the Process and Computing the Coefficients: Having computed the first four conditional moments, we now introduce a scaled process $\bar{Y}_\Delta = \frac{Y_\Delta \mu_1}{\mu_2 - \mu_1^2}$ and set $D = \mu_1^2/(\mu_2 - \mu_1^2) - 1$. Note that

$$\mathbb{E}[\bar{Y}_\Delta \mid Y_0 = y_0, \varrho_Y] = \mathbb{V}[\bar{Y}_\Delta \mid Y_0 = y_0, \varrho_Y] = D + 1.$$

Hence, in view of Lemma 2.2 we match the first two moments of $\bar{Y}_\Delta$ with the ones of the standardized gamma density $w = \gamma(\xi; D)$ from eq. (5.1), since expectation and variance of $\gamma(\xi; D)$ equal $D + 1$ as well.

By using the corresponding orthogonal polynomials $\widetilde{H}_n^\gamma$ from (5.4) along with their normalization constant HO_n^γ from eq. (5.5) we obtain the coefficients of the density approximation

(2.1): for each $n \geq 0$ we have

$$
(7.2) \qquad c_n(y_0, \Delta, \varrho_Y) = \frac{\mathbb{E}\left[\widetilde{H}_n^\gamma(\bar{Y}_\Delta) \mid Y_0 = y_0, \varrho_Y\right]}{HO_n^\gamma}.
$$

In particular, the first five coefficients are of the following explicit form:

$$
c_0(y_0, \Delta, \varrho_Y) = 1
$$
$$
c_1(y_0, \Delta, \varrho_Y) = 0
$$
$$
c_2(y_0, \Delta, \varrho_Y) = 0
$$
$$
c_3(y_0, \Delta, \varrho_Y) = \frac{(D+1)\left((D+2)(D+3) - \frac{(D+1)^2\mu_3}{\mu_1^3}\right)}{\sqrt{6}\sqrt{(D+1)(D+2)(D+3)}}
$$
$$
c_4(y_0, \Delta, \varrho_Y) = \frac{(D+1)^4\mu_4 + (D+4)(D+1)\mu_1\left(3(D+2)(D+3)\mu_1^3 - 4(D+1)^2\mu_3\right)}{2\sqrt{6}\sqrt{(D+1)(D+2)(D+3)(D+4)}\mu_1^4}.
$$

Note that due to the chosen scaling, the deforming polynomial (the pseudo likelihood ratio) does not contribute to the density approximation for the first two orders as predicted by Lemma 2.2. Step 3. Verification of Assumption 2: We denote by g and $\bar{g}$ the density of Y_Δ and $\bar{Y}_\Delta$, respectively. The existence of g and therefore of $\bar{g}$ is ensured by requiring

$$
\kappa\theta > \frac{\sigma^2}{2} > 0,
$$

by Theorem 4.1. By the same result, g and $\bar{g}$ are of class C^p for the greatest nonnegative integer p satisfying

$$
(7.3) \qquad p < \frac{2\kappa\theta}{\sigma^2} - 1.
$$

On the other hand, using Theorem 4.2 one can verify numerically, by solving the corresponding Riccati differential equations, that

$$
(7.4) \qquad \mathbb{E}[e^{\bar{Y}_\Delta}] = \mathbb{E}[e^{\frac{(D+1)}{\mu_1}Y_\Delta}] < \infty.
$$

This implies finite polynomial moments of $\bar{g}$ and g, and therefore justifies the calculations in Steps 1 and 2, in particular. Note that the Gamma density $w(\xi) = \gamma(\xi; D)$ satisfies $\sup_{x \in [0,1]} x^D/w(x) < \infty$ and $\sup_{x \geq 1} e^{-x}/w(x) < \infty$. In view of (7.4), Lemma 3.4 implies validity of Assumption 2, that is $\bar{g}/w \in \mathcal{L}_w^2$, once $D \leq 2p$. By (7.3), the latter holds if and only if[13]

$$
\lceil D/2 \rceil < \frac{2\kappa\theta}{\sigma^2} - 1,
$$

which again can easily be checked numerically.

---

[13]$\lceil x \rceil$ denotes the smallest integer which is greater than or equal to x.

Step 4. Putting Everything Together: Accounting for the change of variable $\bar{y}(y) = \frac{y\mu_1}{\mu_2 - \mu_1^2}$, the density proxy equals

$$(7.5) \qquad g_Y^{(4)}(y \mid y_0, \Delta, \varrho_Y) = \gamma(\bar{y}(y)) \cdot \sum_{i=0}^{4} c_i(y_0, \Delta, \varrho_Y) H_i^\gamma(\bar{y}(y)) \cdot \frac{\mu_1}{\mu_2 - \mu_1^2}.$$

Figure 3 shows how the polynomials $c_i(y_0, \Delta, \varrho_Y) H_i^\gamma(\bar{y}(y))$ in the pseudo likelihood ratio deform the auxiliary density $w = \gamma$ into the right shape.

7.2. Heston's Model.

The Heston (1993) stochastic variance model has been particularly used for the pricing of equity (index) options. The model for the log stock price X and its stochastic variance V can be realized as solution of the following SDE

$$(7.6) \qquad \begin{aligned} dV_t &= (\kappa\theta_V - \kappa_V V_t)\, dt + \sigma\sqrt{V_t}dW_t^V \\ dX_t &= (\kappa\theta_X - \frac{1}{2}V_t)dt + \sqrt{V_t}\left(\rho dW_t^V + \sqrt{1-\rho^2}dW_t^X\right), \end{aligned}$$

with (W^V, W^X) being a two-dimensional standard Brownian motion. The domain $\mathcal{D}$ of the process equals $\mathbb{R}_+ \times \mathbb{R}$. With $2\kappa\theta_V > \sigma^2$ and $|\rho| < 1$ Theorem 4.1 guarantees existence of transition densities. Note that it would be perfectly possible to enrich Heston's model above with jumps in both factors (this has been done for example in Duffie et al. (2000), Eraker et al. (2003), and Eraker (2004)), to multiple variance factors, or even a matrix-valued variance process as in (Da Fonseca et al., 2008).

The correlation parameter ρ above is a device to model the leverage effect which is partly responsible for the skew in option prices. Figure 1 displays, using an order 4 expansion, how the skew of the density may be altered by decreasing the ρ parameter.

For the bivariate Heston model, to compute conditional moments up to order two using formula (4.4), the canonical basis is given by $\{1, v, x, v^2, vx, x^2\}$ and the corresponding Q matrix, analogous to (4.3), is

$$Q = \begin{pmatrix} 0 & \kappa\theta_V & \kappa\theta_X & 0 & 0 & 0 \\ 0 & -\kappa_V & -\frac{1}{2} & \sigma^2 + 2\kappa\theta_V & \kappa\theta_X + \rho\sigma & 1 \\ 0 & 0 & 0 & 0 & \kappa\theta_V & 2\kappa\theta_X \\ 0 & 0 & 0 & -2\kappa_V & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & -\kappa_V & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Example 7.2 (Standardizing and Scaling the Heston Model). Our goal is to work within a $\mathcal{L}^2$ space weighted with a product measure $w : \mathbb{R}_+ \times \mathbb{R} \mapsto \mathbb{R}_+$

$$w(\xi, \eta)d(\xi, \eta) = \gamma(\xi)d\xi \cdot \gamma_b(\eta)d\eta,$$

composed of $\Gamma(D+1, 1)$ and $\Gamma_b(C)$ densities, from definitions (5.1) and (5.2), respectively. In this space, we may use the polynomials from eqs. (5.4) and (5.6). Lemma 3.2 ensures that the product of the polynomials forms an ONB in the $\mathcal{L}_w^2$ space. Acknowledging Lemma 2.2 we want to make sure that we match as many moments as possible to optimize the quality of the approximation. Below we show how this transformation ς is composed.
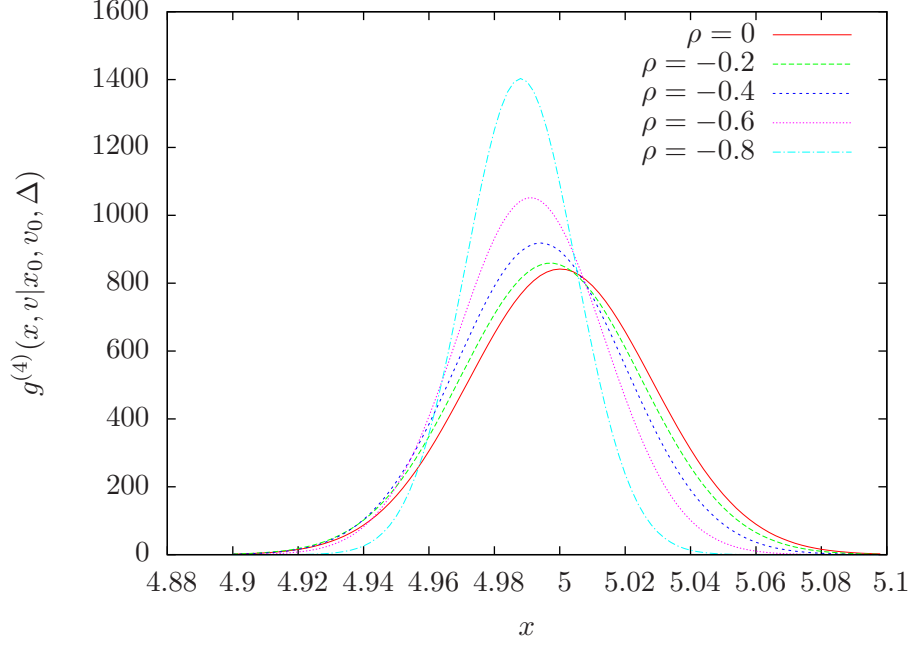
FIGURE 1. **The Effect of Leverage:** The figure shows the effect on skewness of negative correlation between the log stock price and its stochastic variance. Depicted are approximated transition densities $g_{VX}^{(4)}$ of the Heston model for different values of the correlation parameter ρ for fixed $v = 0.043$. The parameters that generated the picture were $\Delta = 1/52, \kappa\theta_V = 0.04, \kappa_V = 1, \sigma = 0.2, \kappa\theta_X = 0.03, x_0 = 5, v_0 = 0.04$.

We introduce lighter notation by defining $U_t = (V_t, X_t)$ and for the first two moments of (V_t, X_t)

$$\mathbb{E}\left[U_t \mid U_0 = u, \varrho_{VX}\right] = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

and

$$\mathbb{V}\left[U_t \mid U_0 = u, \varrho_{VX}\right] = \begin{pmatrix} a_1 & b \\ b & a_2 \end{pmatrix}.$$

For demeaning and block-diagonalizing define

$$\varsigma_1(u) = \Upsilon_1 u + v_1$$

where

$$\Upsilon_1 = \begin{pmatrix} 1 & 0 \\ -b/a_1 & 1 \end{pmatrix}, \quad v_1 = \begin{pmatrix} 0 \\ b\mu_1/a_1 - \mu_2 \end{pmatrix}.$$

Then $\varsigma_1(U_t)$ has first two moments of the form

$$\mathbb{E}\left[\varsigma_1(U_t) \mid U_0 = u, \varrho_{VX}\right] = \begin{pmatrix} \mu_1 \\ 0 \end{pmatrix},$$

$$\mathbb{V}\left[\varsigma_1(U_t) \mid U_0 = u, \varrho_{VX}\right] = \begin{pmatrix} a_1 & 0 \\ 0 & -b^2/a_1 + a_2 \end{pmatrix}.$$

The next transformation scales the process into the optimal form (according to Lemma 2.2)

$$\varsigma_2(u) = \Upsilon_2 \cdot u,$$

where

$$\Upsilon_2 = \begin{pmatrix} \mu_1/a_1 & 0 \\ 0 & 1/\sqrt{-b^2/a_1 + a_2} \end{pmatrix}.$$

Then $\varsigma_2 \circ \varsigma_1(U_t)$ has first two moments of the form

$$\mathbb{E}\left[\varsigma_2 \circ \varsigma_1(U_t) \mid U_0 = u, \varrho_{VX}\right] = \begin{pmatrix} \mu_1^2/a_1 \\ 0 \end{pmatrix},$$

$$\mathbb{V}\left[\varsigma_2 \circ \varsigma_1(U_t) \mid U_0 = u, \varrho_{VX}\right] = \begin{pmatrix} \mu_1^2/a_1 & 0 \\ 0 & 1 \end{pmatrix},$$

and choosing $D = \mu_1^2/a_1 - 1$ the bivariate orthogonal expansion of the density of $\varsigma_2 \circ \varsigma_2(U_t)$ may be performed in terms of the polynomials introduced in (5.4) and (5.6). By the transformation the polynomial moments up to second order induced by w agree with the moments of $\varsigma_2 \circ \varsigma_1(U_t)$ and the moment-matching Lemma 2.2 applies up to order 2. We have used

$$\varsigma : u \mapsto \Upsilon_2 \circ (\Upsilon_1 u - \upsilon_1),$$

and its inverse is

$$\varsigma^{-1} : u \mapsto \Upsilon_1^{-1}\Upsilon_2^{-1}u + \Upsilon_1^{-1}\upsilon_1.$$

The parameter C in (5.2) is set to the exact excess kurtosis of the transformed log stock process and the expansion may be performed analogously to Example 7.1.

7.3. **CDO Pricing.** In the reduced-form credit risk framework (Lando, 1998), we model the stochastic default intensity λ of a corporation with a positive process such as (7.1). Under the pricing measure $\mathbb{Q}$ the default time τ of a corporation is then taken to be the first jump of an inhomogeneous Poisson process with intensity λ. More formally we write the survival probability of a corporation (using the short-hand notation $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$)

$$\mathbb{Q}[\tau > T \mid \mathcal{F}_t] = \mathbb{1}_{\{\tau>t\}}\mathbb{E}_t\left[e^{-\int_t^T \lambda_u du}\right].$$

All expectations are with respect to the risk-neutral pricing measure $\mathbb{Q}$. For the pricing of portfolio credit derivatives, to introduce dependence between different obligors, Duffie and Garleanu (2001) (and subsequently Mortensen (2006), Eckner (2009), and Feldhütter (2008)) introduce a factor intensity model

(7.7) $$\lambda_{it} = X_{it} + a_i Y_t,$$

where X_{it} is a firm-specific (idiosyncratic) intensity factor, and Y_t is a (systemic) factor common to all obligors $i = 1, \ldots, n$. We model both X and Y with independent jump-diffusion processes from eq. (7.1). For n obligors we must impose $\sum_{i=1}^n a_i = 1$ to ensure identifiability (see Eckner (2009)).

The survival probability of obligor i according to model (7.7) is then due to independence of the factors

(7.8) $$\mathbb{Q}[\tau_i > T \mid \mathcal{F}_t] = \mathbb{1}_{\{\tau_i>t\}}\mathbb{E}_t\left[e^{-\int_t^T X_{iu} du}\right]\mathbb{E}_t\left[e^{-a_i \int_t^T Y_u du}\right].$$

Defining $Z_{t,T} = \int_t^T Y_s \, ds$ we may write the default probability of the ith obligor *conditional* on $Z_{t,T}$ as

$$(7.9) \qquad q_i(Z_{t,T}) = \mathbb{Q}_t \left[t < \tau_i \leq T \mid Z_{t,T} \right] = \mathbb{1}_{\{\tau_i > t\}} \left(1 - \mathbb{E}_t \left[e^{-\int_t^T X_{iu} \, du} \right] e^{-a_i Z_{t,T}} \right),$$

and denoting with $P_{t,T}^n(k \mid Z_{t,T})$ the conditional probability that k of the first n credits in the portfolio default between t and T the recursive algorithm of Andersen et al. (2003) then develops the number k of defaults conditional on $Z_{t,T}$ as

$$
(7.10) \qquad
\begin{aligned}
P_{t,T}^{(0)}(k \mid Z_{t,T}) &= \mathbb{1}_{\{k=0\}} \\
P_{t,T}^{(m+1)}(k \mid Z_{t,T}) &= q_{m+1}(Z_{t,T}) P_{t,T}^{(m)}(k-1 \mid Z_{t,T}) + (1 - q_{m+1}(Z_{t,T})) P_{t,T}^{(m)}(k \mid Z_{t,T}),
\end{aligned}
$$

for $0 \leq k \leq n$ and $0 \leq m < n$. The expressions $\mathbb{E}_t \left[e^{-\int_t^T X_{iu} \, du} \right]$, $i = 1, \ldots, n$ are unproblematic, but computing the *unconditional* default probability

$$P_{t,T}^{(n)}(k) = \int P_{t,T}^{(n)}(k \mid Z_{t,T}) d\mathbb{Q}(Z_{t,T})$$

involves an integration against the density of $Z_{t,T}$. We can get hold of the distribution of $Z_{t,T}$ by investigating the joint evolution of Y from eq. (7.1) and the integral over Y. We therefore embed Y into the two-dimensional affine process (Y, Z) described by

$$
(7.11) \qquad
\begin{aligned}
dY &= (\kappa\theta - \kappa Y_t) \, dt + \sigma \sqrt{Y_t} dW_t + dK_t \\
dZ_t &= Y_t dt.
\end{aligned}
$$

Note that even though the instantaneous covariance matrix of the process (7.11) above is only of rank one, this process is a well-defined affine process in the sense of Duffie et al. (2003) as pointed out also in Section 4.3. Existence of the marginal transition density of $Z_{t,T} \mid Y_t$ is shown in Theorem 4.3 for $\kappa\theta > \sigma^2$.

In principle the conditional default probabilities from eq. (7.10) may be computed using the moment generating function of $Z_{t,T}$. In real world applications n is typically larger than 100, however, and the expressions become intractably large, even for small k. In practice, recursion (7.10) is therefore computed through numerical integration. A test of our density expansion in this setting may therefore be reduced to the question of how well we can approximate the true moment generating function. Below we outline how this approximation can be done in closed form.

Denote by $\mathbb{E}_t^{(J)}[f(Z_{t,T})]$ the expectation of $f(Z_{t,T})$ with respect to a J-order expansion instead of the true density. Considering the functional form of the expansion (2.1), to approximate the expressions $\mathbb{E}_t \left[e^{-Z_{t,T} a_i} \right]$, $i = 1, \ldots, n$ we note that we need to perform the computation

$$(7.12) \qquad \mathbb{E}_t^{(J)} \left[e^{a Z_{t,T}} \right] = \int_{\mathbb{R}_+} e^{a\xi} w(\xi) \sum_{j=0}^{J} c_j H_j(\xi) \, d\xi$$

$$(7.13) \qquad \qquad\qquad\quad = \sum_{j=0}^{J} c_H(j) \int_{\mathbb{R}_+} e^{a\xi} \xi^j w(\xi) \, d\xi,$$
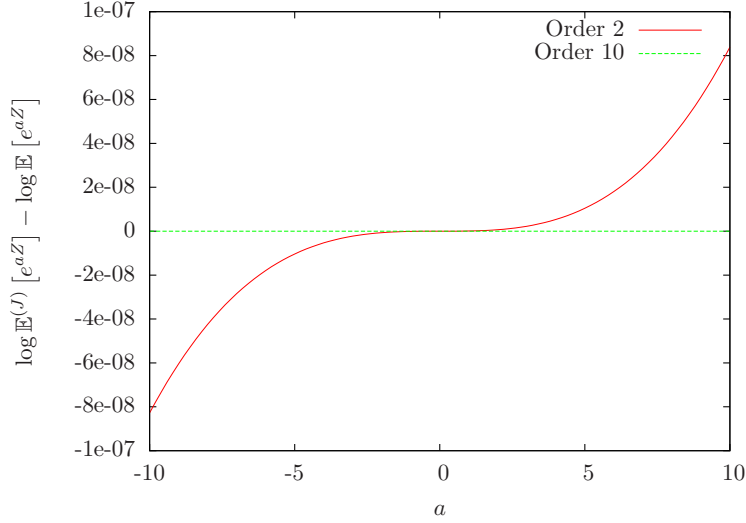
FIGURE 2. **True vs. Approximated Moment Generating Function:** The figure shows the log difference between the true moment generating function $\mathbb{E}_t\left[e^{aZ_{t,T}}\right]$ and the approximated moment generating function $\mathbb{E}_t^{(J)}\left[e^{aZ_{t,T}}\right]$ computed for an order 2 and an order 10 expansion of the integrated BAJD from (7.11). The parameters that generated the picture were $a = 1, T - t = 5, \kappa\theta = 0.00150602, \kappa = 0.4648, \sigma = 0.01, l = 1, \nu = 0.0002, y_0 = (\kappa\theta + l\nu)/\kappa$. Results are computed using Mathematica and the picture is generated with a numeric precision of 20 digits.

where $c_H(j)$ is implicitly defined as[14]

$$\sum_{j=0}^{J} c_j H_j(\xi) = \sum_{j=0}^{J} c_H(j)\xi^j.$$

The chosen $\mathcal{L}^2$ weight w for the approximating transition density is a Gamma distribution. To compute eq. (7.13) for a random variable Z that is Gamma distributed $Z \sim \Gamma(\alpha, \theta)$ we note that
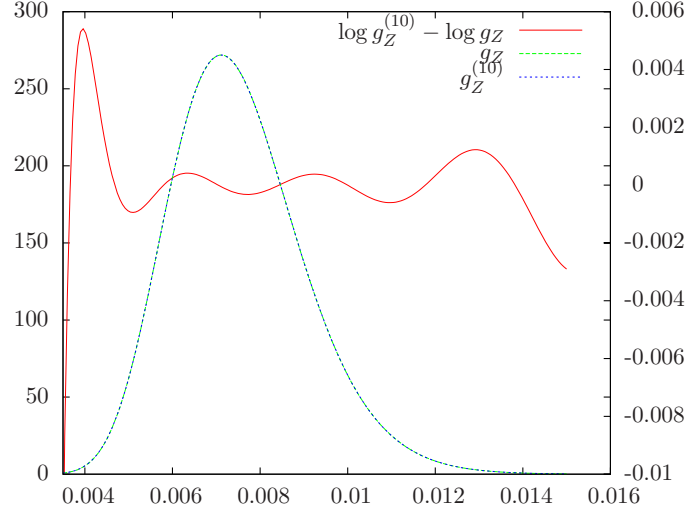
$$\mathbb{E}\left[e^{aZ}Z^n\right] = \frac{\theta^{-n}\Gamma(\alpha - n)(1 - a\theta)^{n-\alpha}}{\Gamma(\alpha)}, \; n \in \mathbb{N}, a \in \mathbb{R},$$
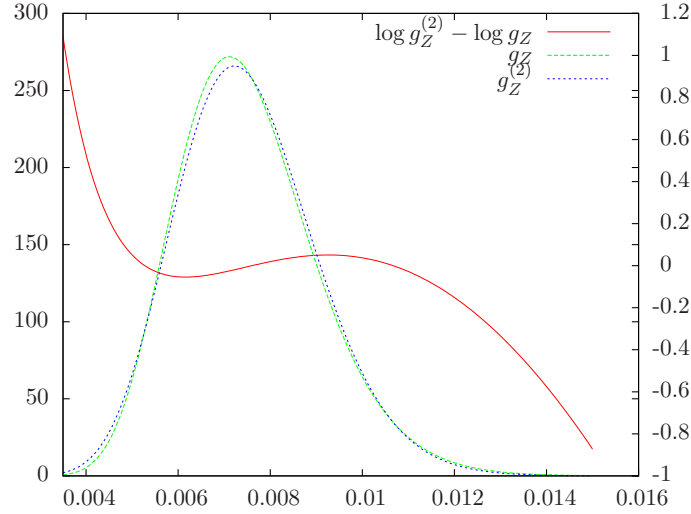
where Γ denotes the Gamma function.

Figure 2 shows that for the order 10 expansion the approximation error is numerically zero. The order 2 expansion also works well, with negligible numeric error.

7.4. **Likelihood-based Inference.** In this section we investigate the performance of the polynomial density expansions in likelihood-based inference. For discrete, equally spaced (with time interval Δ) observations $(X_0, X_1, \ldots, X_N) = X$ of a Markov process $(X_t)_{t\geq 0, X_0 = x_0}$ with domain

---

[14]In practice the coefficients may be collected using a symbolic mathematics package such as Mathematica or Maple.

(A) Order 10 Expansion

(B) Order 2 Expansion

FIGURE 3. **Density Plots of the integrated BAJD:** The figure shows the density of $Z_{0,\Delta} \mid y_0, \varrho_Y$ from specification (7.11). The parameters generating this density are: $\Delta = 5, \kappa\theta = 0.00150602, \kappa = 0.4648, \sigma = 0.01, l = 1, \nu = 0.0002, y_0 = (\kappa\theta + l\nu)/\kappa$. The right y axis shows the deviation error to the true density (obtained by Fourier inversion) in percentage terms.

$\mathcal{D}$ and parameters ϱ_X we may write the likelihood function $l_X : \mathcal{D}^N \times \varrho_X \to \mathbb{R}_+$ as

$$(7.14) \qquad l_X(X \mid \varrho_X, X_0) = \prod_{i=1}^{N} g_X(X_i \mid X_{i-1}, \varrho_X, \Delta).$$

Denote by

$$(7.15) \qquad l_X^{(J)}(X \mid \varrho_X, X_0) = \prod_{i=1}^{N} g_X^{(J)}(X_i \mid X_{i-1}, \varrho_X, \Delta)$$

the approximate likelihood function using a J order expansion developed in this paper. The maximum likelihood estimator $\widehat{\varrho}_X$ is obtained as the global maximizer of the likelihood (7.14)

$$(7.16) \qquad \widehat{\varrho}_X = \arg\max_{\varrho_X} \prod_{i=1}^{N} g_X(X_i \mid X_{i-1}, \varrho_X, \Delta).$$

Similarly, for a maximizer obtained from the approximate likelihood $l_X^{(J)}$ we write

$$(7.17) \qquad \widehat{\varrho}_X^{(J)} = \arg\max_{\varrho_X} \prod_{i=1}^{N} g_X^{(J)}(X_i \mid X_{i-1}, \varrho_X, \Delta).$$

The Bayesian framework (cf. Robert, 1994, for reference and comparison to other methodologies) views the parameters themselves as random variables and is aimed at the posterior density

$$(7.18) \qquad \begin{aligned} p(\varrho_X \mid X) &= \frac{l_X(X \mid \varrho_X)}{\int l(X \mid \varrho)\pi(\varrho)d\varrho}\pi(\varrho_X) \\ &\propto l_X(X \mid \varrho_X)\,\pi(\varrho_X). \end{aligned}$$

The prior density $\pi : \varrho_X \to \mathbb{R}_+$ expresses the econometrician's personal beliefs and knowledge. Its specification may be fueled by economic intuition, for example that nominal interest rates should be positive, and also parameter constraints. Note that the expression (7.18) invokes Bayes theorem and therefore demands that l_X and π actually are densities in that they are non-negative functions on the domain of the random variable and integrate to one. This requirement has been challenged to a great extent. The most common violation stems from expressing uninformedness by setting the prior for a parameter ϱ_+ with positive domain proportional to a constant

$$(7.19) \qquad \pi(\varrho_+) \propto \frac{1}{\sigma_{\varrho_+}}.$$

Prior specifications such as the one mentioned above are called *improper* priors, because their integral does not exist. A less common violation arises from the use of closed-form likelihood expansions within Bayesian inference for Markov processes.[15] For the univariate likelihood expansions from Aït-Sahalia (2002) the normalization constant may be evaluated through numeric integration, putting a heavy computational burden on the econometrician. For the multivariate expansions for irreducible models from Aït-Sahalia (2008) the normalization constant does not even exist, because the expansions are purely polynomial. In contrast, the expansions developed in the present paper integrate to one by construction. They share with the expansions from Aït-Sahalia (2002) the unpleasant feature that they may become negative, however, even though experience shows that this happens very rarely. For the empirical studies in this paper, for instance, it has not happened even once.

---

[15]See Di Pietro (2001) for an introduction to the problem and Stramer et al. (2009) for MCMC algorithms to overcome it in a very general context.

Subsequently we will denote posteriors where the likelihood is approximated using the approximate likelihood $l_X^{(J)}$ from (7.15) by

$$p^{(J)}(\varrho_X \mid X) = l_X^{(J)}(X \mid \varrho_X) \pi(\varrho_X).$$

To test both methodologies we generate realizations from models (7.1) and (7.6) through exact simulation methods. We then perform both frequentist and Bayesian inference using our density approximations and the true density (obtained through Fourier inversion of the characteristic function). Frequentist inference is performed on 1,000 data sets generated by model (7.6), to acquire information about the sampling distribution of the (approximate) maximum likelihood estimators. Bayesian inference is performed on one data set, for the BAJD (eq. (7.1)) and Heston's model (eq. (7.6)), respectively. We then compare the posterior distribution originating from true density to the posterior distribution generated by the density approximations from this paper.

The simulation for each data set is started from the unconditional mean and then propagated forward 600 data points. We discard the first 100 observations to eliminate impact of the initial condition. To investigate the behavior of our density expansions for different time horizons we choose a monthly observation frequency for the square-root jump-diffusion (7.1) and weekly observation frequency for the Heston model (7.6).

To obtain exact draws from the BAJD we generate exact draws from $Y_i \mid Y_{i-1}$ using Robert and Casella (2004, Lemma 2.4). For a uniform random variable $U \sim \mathcal{U}(0,1)$ we exploit that $G_Y^{-1}(U \mid Y_{i-1}, \varrho_Y) \sim G_Y$ for any distribution function G_Y. We simulate from (7.1) using the parameters $\kappa\theta = 0.04, \kappa = 1, \sigma = 0.2, l = 3, \nu = 0.01$.

Algorithm 7.1 (Exact draws from BAJD process (7.1)). We perform the following procedure starting from $Y_0 = \mathbb{E}[Y_t]$, the unconditional mean, for a realization $Y_i \mid Y_{i-1}$

    (i) Draw $U \sim \mathcal{U}(0,1)$. Call the realization u_i

    (ii) Use the Newton-Raphson algorithm to compute $y : G_Y(Y \leq y \mid Y_{i-1}, \varrho_Y) = u_i$. In this step we substitute $y = \frac{e^w}{1+e^w} + c$ to keep y on the positive domain. The floor parameter c we set to 10^{-6} to avoid numerical difficulties. The iteration is then

$$w_{j+1} = w_j - e^{-w_j}\left(e^{w_j} + 1\right)^2 \frac{G_Y(Y \leq c + \frac{e^{w_j}}{e^{w_j}+1} \mid Y_{i-1}, \varrho_Y) - u_i}{g_Y(c + \frac{e^{w_j}}{e^{w_j}+1} \mid Y_{i-1}, \varrho_Y)}$$

starting from $w_0 = \log\left(\frac{y_{i-1}-c}{c-y_{i-1}+1}\right)$. Stop the iteration at

$$w^\star : \left| G_Y(Y \leq c + \frac{e^{w^\star}}{e^{w^\star} + 1} \mid Y_{i-1}, \varrho_Y) - u_i \right| < \varepsilon.$$

Both, g_Y, and, G_Y are obtained through Fourier inversion. In our implementation the algorithm terminates after 5 to 6 iterations for $\varepsilon = 10^{-6}$.

    (iii) Set $Y_i = c + \frac{e^{w^\star}}{e^{w^\star}+1}$ increment i and go back to step (1)

For Bayesian inference we specify an uninformative prior

(7.20) $$\pi(\varrho_Y) = \mathbb{1}_{\{2\kappa\theta > \sigma^2, \sigma > 0, l > 0, \nu > 0\}} \frac{1}{\sigma \cdot \kappa\theta \cdot l \cdot \nu}.$$

The Heston parameters are $\kappa_V = 1, \kappa\theta_V = 0.04, \sigma = 0.2, \kappa\theta_X = 0.03, \rho = -0.8$. To obtain exact draws from this model we refer the reader to the algorithm in Broadie and Kaya (2006). For Bayesian inference we specify the prior distribution as

$$(7.21) \qquad \pi(\varrho_{VX}) = \mathbb{1}_{\{2\kappa\theta_V > \sigma^2, -1 < \rho < 1, \sigma > 0\}} \frac{1}{\sigma \cdot \kappa\theta_V}.$$

To evaluate the transition density we employ the formulation from Lamoureux and Paseka (2005), which may be evaluated using a single numerical integral, instead of the two-dimensional Fourier integral. This reduction of dimensionality comes at the price of having to evaluate complex-valued special functions, however.

With 1,000 datasets of weekly realizations from the Heston model, for each dataset we obtain parameters $\varrho_{VX}^\star$ by maximizing the log likelihood (7.14), respectively the approximate log likelihood (7.15). We use the optimizer `donlp2` to achieve this task. To relate the density expansions of this paper to existing approximations we perform the estimation experiment with

- the true density (obtained through Fourier inversion) denoted by MLE
- order 4 likelihood expansions developed in this paper using a product measure with a Gamma weight for the variance process and for the log stock variable a
  - bilateral Gamma weight. Specifically we employ formulation (5.3). Estimates are denoted by $BG(4)$
  - Gaussian weight. Estimates are denoted by $G(4)$
- order 2 closed-form likelihood expansions from Aït-Sahalia (2008) denoted by $CF(2)$
- Gaussian approximation using true conditional moments up to order 2 denoted by QML

Table 3 reveals that the true likelihood function exhibits problematic behavior for some parameterizations. Only 688 out of 1,000 estimates turned out to be successful. This is due to numerical integration problems that occur in particular for low values of σ that arise in the likelihood search. Density expansions developed in this paper are also not entirely unproblematic. Numerical errors from evaluating the pseudo likelihood ratio accumulate and induce spikes that irritate the optimizer's numerical differentiation routines. The Hermite polynomials used for $G(4)$ appear better behaved than the polynomials associated with the bilateral Gamma density used in $BG(4)$.

Table 3a reports bias and RMSE of the estimators. The large bias of 0.2255 for the κ parameter is a well-established phenomenon that has also been reported in Aït-Sahalia and Kimmel (2007). As an overall impression the results suggest that the density approximations developed in this paper exhibit parameter estimates with properties similar to the true ML estimates, while Aït-Sahalia (2008) expansions interestingly exhibit lower bias, with the exception of the σ parameter, but higher RMSE. In Table 3b the first column (Mean) in $\widehat\varrho_{VX}^{\star MLE} - \varrho_{VX}^{TRUE}$ indicates mean deviation from the true ML estimator and the second column (SD) captures statistical noise in the estimation. Estimation bias around the MLE for all estimators appears very small. Except for the $CF(2)$ estimator, the noise induced through the density approximations is smaller than the estimation noise of the true MLE. Surprisingly, the QML estimator, a special case of the approximations developed in this paper since it is an order two expansion around a Gaussian, performs remarkably well. All around the BG(4) expansions appear to be the preferable choice. In particular $\widehat\sigma^{\star BG(4)} - \widehat\sigma^{\star MLE}$ and $\widehat\rho^{\star BG(4)} - \widehat\rho^{\star MLE}$ point to the right direction, the estimators are closer to the true parameters than MLE.

The results of the Bayesian inference study also appear promising. Inspecting Figures 5 we see that an order 2 expansion already delivers reasonable results, while the order 4 expansion seems to be even closer to the posterior density obtained from the true density function. To assess how close the posteriors $p_{VX}^{(2)}$ and $p_{VX}^{(4)}$ densities are to the posterior obtained through the true transition density p_{VX} we compute Kolmogorov-Smirnov tests. The results can be seen in Table 2. They suggest that while $p_{VX}^{(2)}$ appears to be quite different from p_{VX}, $p_{VX}^{(4)}$ is statistically almost indistinguishable from the true posterior p_{VX} for the majority of the parameters.

7.5. **Option Pricing.** Heston's model (7.6) is used for option pricing because it may be consistent with the implied volatility skew that can be inferred from market prices. As such it is much more compatible with real data than say, the Black-Scholes model. In stock (index) option pricing the quantity of interest are marginal transition probabilities of the log stock price X. We therefore engineer an approximation directly around the marginal density of $X_\Delta \mid X_0, V_0$ by expanding g_X in $\mathcal{L}^2_{\gamma_b}$. We set the constant C from (5.2) to the excess kurtosis of X.

Recall that the price of a European call option with maturity Δ and strike price K is given by

$$C(\Delta, K) = e^{-r\Delta} \mathbb{E}\left[\left(e^{X_\Delta} - K \right)^+ | X_0 = x, V_0 = v, \varrho_{VX} \right]$$

(7.22)
$$= e^{-r\Delta} \left(\underbrace{\int_{\log K}^\infty e^\xi g_X(\xi | x, v, \varrho_{VX}, \Delta) d\xi}_{HA(\Delta, K)} - K \underbrace{\int_{\log K}^\infty g_X(\xi | x, v, \varrho_{VX}, \Delta) d\xi}_{HB(\Delta, K)} \right).$$

In accordance with the previous sections we will denote $C^{(J)}(\Delta, K)$, and similarly $HA^{(J)}(\Delta, K)$ and $HB^{(J)}(\Delta, K)$, the option price computed with $g_X^{(J)}$ instead of g_X. Denoting $\mathbb{Q}(X \le x)$ the transition probability (and accordingly $\mathbb{Q}^{(J)}(X \le x)$ the J order approximation of the transition probability) we have that $HB^{(J)}(\Delta, K) = 1 - \mathbb{Q}^{(J)}(X \le \log K)$, and using the standardization from Example 7.2 and the change of variables formula

$$HB^{(J)}(\Delta, K) = 1 - \int_{-\infty}^{\log K} g_X^{(J)}(\xi | x, v, \varrho_{VX}) d\xi$$

(7.23)
$$= 1 - \frac{1}{\sqrt{a_2}} \int_{-\infty}^{\log K} \gamma_b \left(\frac{\xi - \mu_2}{\sqrt{a_2}} \right) \left(1 + \sum_{i=1}^J c_i(x, v, \varrho_{VX}) H_i^{\gamma_b} \left(\frac{\xi - \mu_2}{\sqrt{a_2}} \right) \right) d\xi$$

$$= 1 - \frac{1}{\sqrt{a_2}} \sum_{i=0}^J \Gamma_b \left(\frac{\log K - \mu_2}{\sqrt{a_2}}, i \right) \vartheta_i(x, v, \varrho_{VX}).$$

Here, $H_i^{\gamma_b}$ are from eqs. (5.6) and (5.7) and $\vartheta_i(x, v, \varrho_{VX})$ are implicitly defined as

$$\left(1 + \sum_{i=1}^J c_i(x, v, \varrho_{VX}) H_i^{\gamma_b} \left(\frac{\xi - \mu_2}{\sqrt{a_2}} \right) \right) = \sum_{i=0}^J \vartheta_i(x, v, \varrho_{VX}) \xi^i.$$

The function $\Gamma_b(K, n) = \int_{-\infty}^K \xi^n \gamma_b(\xi) d\xi$ is explicit in terms of the Gamma function and regularized generalized hypergeometric functions. The constituent $HA^{(J)}$ of the approximate call price
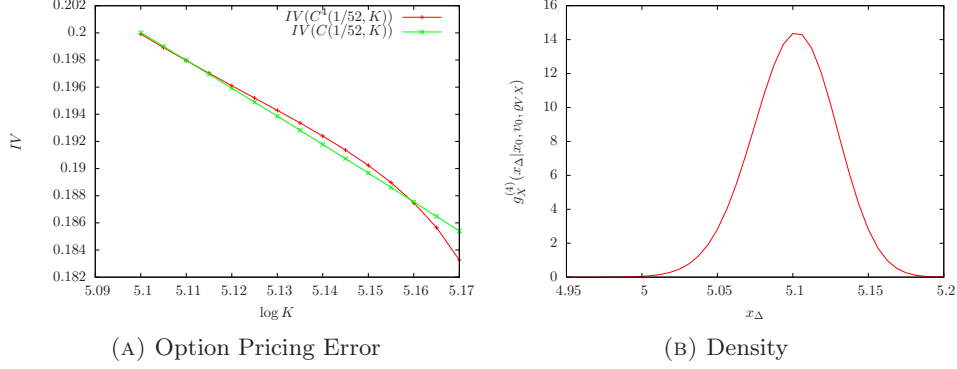
(A) Option Pricing Error   (B) Density

FIGURE 4. **Closed-form option pricing in Heston's model** Panel 4a shows implied Black-Scholes volatility of the true option price, here computed using the Carr and Madan (1999) dampened Fourier inversion approach and the $C^{(4)}(\Delta, K)$ option pricing formula from the approximation of eq. (7.22) as a function of strike K. The second panel 4b shows the density function of $X_\Delta \mid X_0, V_0$ to indicate the likelihood-moneyness trade off. The parameters for the model (7.6) behind the pictures are $\Delta = 1/52, \kappa\theta_V = 0.04, \kappa_V = 1, \sigma = 0.2, \kappa\theta_X = 0.03, \rho = -0.8, X_0 = 5.1, V_0 = 0.04$.

from eq. (7.22) can be computed by numeric integration. Figure 4 shows that option pricing performance is very good.

Remark 7.2. Collecting coefficients to compute ϑ_i from Section 7.5 eq. (7.23) by hand is very error-prone. Instead we recommend using a symbolic mathematics software package such as Mathematica or Maple.

8. DISCUSSION

This paper develops a general framework for density approximations for affine processes using orthonormal polynomial expansions in well-chosen weighted $\mathcal{L}^2$ spaces. We provide novel existence and smoothness results for their transition densities in particular.

The approximations are designed to exploit the explicit polynomial moments of affine processes to compute the coefficients of the expansion without approximation error and in closed form; the computational burden is concentrated only in the initial calculation of the coefficients of the expansions. Once they are implemented, evaluation is rapid, avoiding the heavy computational cost of Fourier methods to obtain transition densities. Empirical applications in credit risk, likelihood-based parameter inference, and option pricing suggest that the density expansions are very accurate.

The paper leaves a number of open points for future research. The first question concerns approximations in higher-order weighted Sobolev spaces. One might suspect that approximation of (sufficiently smooth) densities in weighted higher-order Sobolev spaces are superior to $\mathcal{L}^2$ expansions. In particular, it could be expected that (i) the quality of approximation might be better (ii) Sobolev embedding theorems could be applied to infer global uniform convergence. However, quite contrary to the $\mathcal{L}^2$ case, it is unknown whether the space of polynomials is dense in weighted Sobolev spaces. Higher-order Sobolev spaces also impose heavy restrictions on the

functional form of the approximation weights, which in turn lead to very slow convergence rates. Indeed, preliminary numerical experiments suggest that the price for global, uniform convergence which potentially comes with higher-order Sobolev spaces is a very slow convergence rate.

Another route worth pursuing is a compact truncation of the state space, such that approximations could be performed in non-weighted Sobolev spaces, for which there is more theory available in the literature.

Suitable approximation weights (such as the bilateral gamma weight of this paper) are a research topic of its own, and they lead to non-trivial problems in the theory of special functions. Also, density expansions for processes on state spaces different from the canonical ones would be highly desirable. As an example we mention the class of matrix-valued processes used in co-volatility modeling (Leippold and Trojani, 2008; Da Fonseca et al., 2008; Buraschi et al., 2008).

Appendix A. Proofs for Section 3

This appendix gathers the proofs of the lemmas in Section 3.

Proof of Lemma 3.1. That the set of polynomials is dense in $\mathcal{L}^2_w$ is shown in Bernard (1995, Lemma 1). The assumption made in Bernard (1995) that w is strictly positive can easily be omitted by replacing point-wise equality "$= 0$" by "$= 0\ w(\xi)\,d\xi$-a.s." at the end of the proof of Bernard (1995, Lemma 1). An orthonormal basis of polynomials $\{H_\alpha \mid \alpha \in \mathbb{N}_0^d\}$ of $\mathcal{L}^2_w$ with $\deg H_\alpha = |\alpha|$ is obtained by applying the Gram–Schmidt process to the linearly independent set of monomials $\{\xi^\alpha \mid \alpha \in \mathbb{N}_0^d\}$, see Algorithm 5.1. $\qquad\square$

Proof of Lemma 3.2. That the product density w on $\mathbb{R}^d$ has finite exponential moment (3.1) follows from the elementary inequality $\|\xi\| \leq \sum_{i=1}^d |\xi_i|$ for all $\xi \in \mathbb{R}^d$. The orthonormality of H_α follows from the easily verifiable relationship

$$\langle H_\alpha, H_\beta \rangle_{\mathcal{L}^2_w} = \prod_{i=1}^d \left\langle H^i_{\alpha_i}, H^i_{\beta_i} \right\rangle_{\mathcal{L}^2_{w_i}(\mathbb{R})}, \quad \alpha,\,\beta \in \mathbb{N}_0^d.$$

Moreover, every monomial $\xi^\alpha = \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$ can be written as a product of linear combinations of the respective orthonormal polynomials

$$\xi_i^{\alpha_i} = \sum_{j=0}^{\alpha_i} \gamma_{ij} H^i_j(\xi_i).$$

It follows that the set $\{H_\alpha \mid \alpha \in \mathbb{N}_0^d\}$ is dense in $\mathcal{L}^2_w$, and hence forms an orthonormal basis of $\mathcal{L}^2_w$. This proves the lemma. $\qquad\square$

Proof of Lemma 3.3. The lemma follows from the estimate

$$\int_{\mathbb{R}^d} \frac{g(\xi)^2}{w(\xi)}\,d\xi = \int_{\mathbb{R}^d} g(\xi)\,\frac{\mathrm{e}^{-\epsilon_0\|\xi\|}}{w(\xi)}\,\mathrm{e}^{\epsilon_0\|\xi\|}\,g(\xi)\,d\xi$$

$$\leq \sup_{x\in\mathbb{R}^d}\left(g(x)\,\frac{\mathrm{e}^{-\epsilon_0\|x\|}}{w(x)}\right)\int_{\mathbb{R}^d}\mathrm{e}^{\epsilon_0\|\xi\|}\,g(\xi)\,d\xi < \infty.$$

$\qquad\square$

Proof of Lemma 3.4. A Taylor expansion of $g(x)$ around 0 gives $g(x) = \frac{\partial_x^p g(\xi)}{p!} x^p$ for some $\xi = \xi(x) \in [0, x]$. Since $\partial_x^p g$ is continuous we conclude that there exists some finite constant K such that $g(x) \leq K\, x^p$ for all $x \in [0, 1]$. We then obtain

$$\int_0^\infty \frac{g(\xi)^2}{w(\xi)}\, d\xi = \int_0^1 \frac{g(\xi)^2}{w(\xi)}\, d\xi + \int_1^\infty g(\xi)\, \frac{e^{-\epsilon_0 \xi}}{w(\xi)}\, e^{\epsilon_0 \xi}\, g(\xi)\, d\xi$$

$$\leq K^2 \int_0^1 \frac{\xi^{2p}}{w(\xi)}\, d\xi + \sup_{x \geq 1}\left(g(x)\frac{e^{-\epsilon_0 x}}{w(x)}\right)\int_1^\infty e^{\epsilon_0 \xi}\, g(\xi)\, d\xi < \infty.$$

This proves the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Lemma 3.5. Let $x \in \mathcal{I}$, and let i^* be such that $x_{i^*} = \min_i x_i$. A Taylor expansion of $g(x, y)$ in x_{i^*} around $x_{i^*} = 0$ gives $g(x, y) = \frac{\partial_{x_{i^*}}^p g(\xi, y)}{p!} x_{i^*}{}^p$ for some $\xi = \xi(x, y) \in \mathcal{I}$. Since $\partial_{x_i}^p g(x, y)$ is bounded on $\mathcal{I} \times \mathbb{R}^n$ we conclude that there exists some finite constant K such that $g(x, y) \leq K \min_i x_i^p$ for all $(x, y) \in \mathcal{I} \times \mathbb{R}^n$. We then decompose

$$\int_{\mathbb{R}_+^m} \int_{\mathbb{R}^n} \frac{g(\xi, \eta)^2}{w(\xi, \eta)}\, d\xi\, d\eta = I_1 + I_2$$

with

$$I_1 = \int_{\mathcal{I}} \int_{\mathbb{R}^n} \frac{g(\xi, \eta)^2}{w(\xi, \eta)}\, d\xi\, d\eta$$

$$\leq K \int_{\mathcal{I}} \int_{\mathbb{R}^n} \frac{\min_i \xi_i^p\, e^{-\epsilon_2 \|\eta\|}}{w(\xi, \eta)} e^{\epsilon_2 \|\eta\|}\, g(\xi, \eta)\, d\xi\, d\eta$$

$$\leq K \sup_{(x,y)\in\mathcal{I}\times\mathbb{R}^n}\left(\frac{\min_i x_i^p\, e^{-\epsilon_2\|y\|}}{w(x, y)}\right)\int_{\mathcal{I}} \int_{\mathbb{R}^n} e^{\epsilon_2 \|\eta\|}\, g(\xi, \eta)\, d\xi\, d\eta < \infty$$

and

$$I_2 = \int_{(1,\infty)^m} \int_{\mathbb{R}^n} \frac{g(\xi, \eta)^2}{w(\xi, \eta)}\, d\xi\, d\eta$$

$$= \int_{(1,\infty)^m} \int_{\mathbb{R}^n} g(\xi, \eta)\, \frac{e^{-\epsilon_1 \|\xi\| - \epsilon_2 \|\eta\|}}{w(\xi, \eta)}\, e^{\epsilon_1 \|\xi\| + \epsilon_2 \|\eta\|}\, g(\xi, \eta)\, d\xi\, d\eta$$

$$\leq \sup_{(x,y)\in(1,\infty)^m\times\mathbb{R}^n}\left(g(x, y)\frac{e^{-\epsilon_1\|x\|-\epsilon_2\|y\|}}{w(x, y)}\right)\int_{(1,\infty)^m} \int_{\mathbb{R}^n} e^{\epsilon_1 \|\xi\| + \epsilon_2 \|\eta\|}\, g(\xi, \eta)\, d\xi\, d\eta$$

$$< \infty.$$

This proves the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

APPENDIX B. PROOF OF THEOREM 4.1

First we note that the existence and smoothness properties of a density on $\mathbb{R}^d$ for $X_t|X_0 = x$ is invariant with respect to non-singular linear transformations of the state vector X_t. In view of Filipović (2009, Theorem 10.7) there exists a non-singular linear transformation of the state

vector X_t mapping $\mathcal{D} = \mathbb{R}_+^m \times \mathbb{R}^n$ onto itself, and which renders block diagonal matrices α_i in the form

$$(\text{B.1}) \qquad \alpha_i = \begin{pmatrix} \text{diag}(0, \ldots, 0, \alpha_{i,ii}, 0, \ldots, 0) & 0 \\ 0 & \alpha_{i,JJ} \end{pmatrix},$$

so that $x^\top \alpha_i x = \alpha_{i,ii} x_i^2 + x_J^\top \alpha_{i,JJ} x_J$ for all $x \in \mathbb{R}^d$. Moreover, this transformation does not affect the upper diagonal element $\alpha_{i,ii}$ (see the proof of Filipović (2009, Lemma 10.5)), which is important in view of the criterion (4.6). Hence without loss of generality we shall from now on assume that the matrices α_i are of the block diagonal form (B.1).

We now recall a classical result on characteristic functions $\widehat{\nu}$ of probability measures ν, see Sato (1999, Proposition 28.1):

Lemma B.1. *Let ν be a probability measure on $\mathbb{R}^d$. Assume its characteristic function $\widehat{\nu}(iu) = \int_{\mathbb{R}^d} e^{iu^\top \xi} \nu(d\xi)$ satisfies*

$$\int_{\mathbb{R}^d} |\widehat{\nu}(iu)| \, \|u\|^k \, du < \infty$$

for some nonnegative integer k. Then ν has a density $h(x)$ of class C^k and the partial derivatives of $h(x)$ of orders $0, \ldots, k$ tend to 0 as $\|x\| \to \infty$.

It thus remains to prove the appropriate integrability of the affine characteristic function (4.1), that is, the appropriate tail behavior in $u \in \mathbb{R}^d$ of the functions $\phi(t, iu)$ and $\psi(t, iu)$. The following lemma is our core result, which together with Lemma B.1 completes the proof of Theorem 4.1.

Lemma B.2. *The following properties are equivalent:*

(i) *The $d \times (n+1)d$-matrix $\mathcal{K}$ given in (4.5) has full rank.*

(ii) *For any $t > 0$, the $d \times d$-matrix*

$$A(t) = \int_0^t \text{diag}(0, e^{\mathcal{B}_{JJ}^\top s} \, a \, e^{\mathcal{B}_{JJ} s}) ds + t \sum_{i \in L} \alpha_i$$

is nonsingular.

(iii) *For any $t > 0$ there exists an $\epsilon > 0$ such that the cones*

$$C_0 = \left\{ u \in \mathbb{R}^d \mid u_J^\top \left(\int_0^t e^{\mathcal{B}_{JJ}^\top s} \, a \, e^{\mathcal{B}_{JJ} s} \, ds \right) u_J \geq \epsilon \|u\|^2 \right\}$$

$$C_i = \left\{ u \in \mathbb{R}^d \mid u^\top \alpha_i \, u \geq \epsilon \|u\|^2 \right\}, \quad i = 1, \ldots, m$$

cover $\mathbb{R}^d$. That is, $\bigcup_{i=0}^m C_i = \mathbb{R}^d$.

Moreover, any of the above properties, (i), (ii), *or* (iii), *implies that*

$$(\text{B.2}) \qquad \int_{\mathbb{R}^d} \left| e^{\phi(t,iu) + \psi(t,iu)^\top x} \right| \|u\|^p \, du < \infty$$

for all nonnegative numbers $p < \min_{i \in \{1, \ldots, m\}} \frac{b_i}{\alpha_{i,ii}} - 1$.

The remainder of this section is devoted to the proof of Lemma B.2. Let $t > 0$ and $u \in \mathbb{R}^d \backslash \{0\}$. We first claim that $u^\top \mathcal{K} = 0$ if and only if $u^\top A(t) = 0$, which proves equivalence of (i) and (ii). To prove the claim, note that since $A(t)$ is positive semidefinite, $u^\top A(t) = 0$ is equivalent to

$u^\top A(t)u = 0$. Since each of the summands in $A(t)$ is positive semidefinite, this again is equivalent to

$$u^\top \sum_{i=1}^m \alpha_i = 0 \quad \text{and} \quad u_J^\top e^{\mathcal{B}_{JJ}^\top s} a = 0 \quad \text{for all } s \in [0, t].$$

A power series expansion of $e^{\mathcal{B}_{JJ}^\top s}$ shows that this is equivalent to

$$u^\top \sum_{i=1}^m \alpha_i = 0 \quad \text{and} \quad u_J^\top (\mathcal{B}_{JJ}^k)^\top a = 0 \quad \text{for all } k \in \mathbb{N}_0.$$

The Cayley–Hamilton theorem (see (Horn and Johnson, 1990, Theorem 2.4.2)) implies that, for all $k \geq n$, $\mathcal{B}_{JJ}^k$ is a linear combination of Id, $\mathcal{B}_{JJ}, \ldots, \mathcal{B}_{JJ}^{n-1}$. Whence the above property is equivalent to $u^\top \mathcal{K} = 0$, which proves the claim.

The equivalence of (ii) and (iii) follows from the identity

$$u^\top A(t)\, u = u_J^\top \left(\int_0^t e^{\mathcal{B}_{JJ}^\top s}\, a\, e^{\mathcal{B}_{JJ} s}\, ds \right) u_J + t \sum_{i \in L} u^\top \alpha_i\, u$$

and the fact that each of the summands is nonnegative. This establishes the first part of Lemma B.2.

As for the second part of Lemma B.2, we note that as a consequence of (B.1) the real and imaginary parts

$$f(t, iu) = \Re \psi(t, iu) \quad \text{and} \quad g(t, iu) = \Im \psi(t, iu)$$

of ψ satisfy the following system of Riccati equations, for $i = 1, \ldots, m$:

$$\partial_t f_i = \alpha_{i,ii} f_i^2 - g^\top \alpha_i\, g + \mathcal{B}_i\, f + \int_{\mathcal{D}} \left(e^{f^\top \xi} \cos\left(g^\top \xi\right) - 1 \right) \mu_i(d\xi)$$

$$f_i(0) = 0$$

$$f_J \equiv 0$$

$$\partial_t g_i = 2 f_i\, \alpha_{i,ii}\, g_i + \mathcal{B}_i\, g + \int_{\mathcal{D}} e^{f^\top \xi} \sin\left(g^\top \xi\right) \mu_i(d\xi)$$

$$g_i(0) = u_i$$

$$g_J = e^{\mathcal{B}_{JJ} t} u_J$$

In the sequel we will make use, without further notice, of the fact that f_i is $\mathbb{R}_-$-valued for all $i = 1, \ldots, m$, and that $f_j = 0$ for all $j \in J$. In particular, it follows from above that f_i satisfies the following system of differential inequalities

$$\partial_t f_i \leq \alpha_{i,ii}\, f_i^2 - g^\top \alpha_i\, g + \mathcal{B}_{ii}\, f_i, \quad i = 1, \ldots, m.$$

For any $u \neq 0$, we now define the scaled functions

(B.3)
$$F(t, u) = \frac{1}{\|u\|} f\left(\frac{t}{\|u\|}, iu \right)$$
$$G(t, u) = \frac{1}{\|u\|} g\left(\frac{t}{\|u\|}, iu \right).$$

Then F and G satisfy, for $i = 1, \ldots, m$:

$$\partial_t F_i \leq \alpha_{i,ii} \, F_i^2 - G^\top \alpha_i \, G + \frac{1}{\|u\|} \mathcal{B}_{ii} \, F_i$$

$$F_i(0) = 0$$

$$F_J \equiv 0$$

(B.4)
$$\partial_t G_i = 2 F_i \, \alpha_{i,ii} \, G_i + \frac{1}{\|u\|} (\mathcal{B}_i + \mathcal{K}_i) \, G$$

$$G_i(0) = \frac{u_i}{\|u\|}$$

$$G_J = e^{\mathcal{B}_{JJ} t} \frac{u_J}{\|u\|}$$

where we define the $d \times d$-matrix $\mathcal{K} = \mathcal{K}(t, u)$ by its $1 \times d$-row vectors

$$\mathcal{K}_i = \begin{cases} \int_{\mathcal{D}} \left(\int_0^1 \cos \left(s \|u\| G^\top \xi \right) ds \right) e^{\|u\| F^\top \xi} \, \xi^\top \mu_i(d\xi), & i = 1, \ldots, m \\ 0, & i = m+1, \ldots, d, \end{cases}$$

and we have used the simple fact that

$$e^{\|u\| F^\top \xi} \sin \left(\|u\| G^\top \xi \right) = e^{\|u\| F^\top \xi} \int_0^1 \frac{d}{ds} \sin \left(s \|u\| G^\top \xi \right) ds$$

$$= \|u\| G^\top \xi \, e^{\|u\| F^\top \xi} \int_0^1 \cos \left(s \|u\| G^\top \xi \right) ds.$$

It follows by the assumptions on μ_i that the matrix $\mathcal{K}$ is uniformly bounded

$$\sup_{t,u} \|\mathcal{K}(t, u)\| = K < \infty$$

where the constant K only depends on the measures μ_i. The squared norm of G thus satisfies

$$\partial_t \|G\|^2 = 2 G^\top \partial_t G = 4 G^\top \mathrm{diag}(F) \, G + \frac{2}{\|u\|} G^\top (\mathcal{B} + \mathcal{K}) \, G$$

$$\leq \frac{2}{\|u\|} \left(\|\mathcal{B}\| + K \right) \|G\|^2$$

$$\|G(0)\|^2 = 1.$$

We shall now and in the sequel make use of the following comparison result, which is a special case of a more general theorem proved by Volkmann (1972):

Lemma B.3. *Let $R(t, v)$ be a continuous real map on $\mathbb{R}_+ \times \mathbb{R}$ and locally Lipschitz continuous in v. Let $p(t)$ and $q(t)$ be differentiable functions satisfying*

$$\frac{d}{dt} p(t) \leq R(t, p(t))$$

$$\frac{d}{dt} q(t) = R(t, q(t))$$

$$p(0) \leq q(0).$$

Then we have $p(t) \leq q(t)$ for all $t \geq 0$.

Applying Lemma B.3 to the above differential inequality for $\|G\|^2$ we obtain

(B.5) $$\|G\|^2 \leq e^{\frac{2}{\|u\|}(\|\mathcal{B}\|+K)t}.$$

For any $i \in \{1, \ldots, m\}$ we then obtain the differential inequality

$$\partial_t \left(G^\top \alpha_i\, G\right) = 2G^\top \alpha_i\, \partial_t G$$

$$= 4G^\top \alpha_i \operatorname{diag}(F)\, \alpha_i\, G + \frac{2}{\|u\|}\, G^\top \alpha_i (\mathcal{B}+\mathcal{K})\, G$$

$$\geq 4\alpha_{i,ii}\, F_i\, \alpha_{i,ii}\, G_i^2 - \frac{2\|\alpha_i\|}{\|u\|}\left(\|\mathcal{B}\|+K\right)\|G\|^2$$

$$\geq 4\alpha_{i,ii}\, F_i\, G^\top \alpha_i\, G - \frac{2\|\alpha_i\|}{\|u\|}\left(\|\mathcal{B}\|+K\right) e^{\frac{2}{\|u\|}(\|\mathcal{B}\|+K)t}$$

$$G(0)^\top \alpha_i\, G(0) = \frac{u^\top \alpha_i u}{\|u\|^2}$$

where we have used the fact that $\alpha_{i,ii}\, G_i^2 \leq G^\top \alpha_i\, G$ and (B.5) for the last inequality. Lemma B.3 again yields the lower bound

$$G^\top \alpha_i\, G \geq e^{4\alpha_{i,ii} \int_0^t F_i(s)\, ds}\, \frac{u^\top \alpha_i u}{\|u\|^2}$$

$$- \frac{2\|\alpha_i\|}{\|u\|}\left(\|\mathcal{B}\|+K\right) \int_0^t \underbrace{e^{4\alpha_{i,ii} \int_s^t F_i(r)\, dr}}_{\leq 1}\, e^{\frac{2}{\|u\|}(\|\mathcal{B}\|+K)(t-s)}\, ds$$

$$\geq e^{4\alpha_{i,ii} \int_0^t F_i(s)\, ds}\, \frac{u^\top \alpha_i u}{\|u\|^2} - \|\alpha_i\| \left(e^{\frac{2}{\|u\|}(\|\mathcal{B}\|+K)t} - 1 \right)$$

Combining this with the differential inequality (B.4) for F we obtain

(B.6)
$$\partial_t F_i \leq \alpha_{i,ii}\, F_i^2 + \frac{1}{\|u\|}\mathcal{B}_{ii}\, F_i - e^{4\alpha_{i,ii} \int_0^t F_i(s)\, ds}\, \frac{u^\top \alpha_i u}{\|u\|^2}$$
$$+ \|\alpha_i\| \left(e^{\frac{2}{\|u\|}(\|\mathcal{B}\|+K)t} - 1 \right)$$

$$F(0) = 0.$$

We arrive at the following intermediate result.

Lemma B.4. *For every $\epsilon > 0$ and $t_0 > 0$ there exists some $\rho > 0$ and $R > 0$ such that*

$$F_i(t_0, iu) \leq -\rho$$

for all $u \in \mathbb{R}^d$ with $\|u\| \geq R$ and $u^\top \alpha_i u \geq \epsilon \|u\|^2$, for all $i \in \{1, \ldots, m\}$.

Proof. The differential inequality (B.6) is autonomous and smooth in F_i. Moreover, the initial slope satisfies

$$\partial_t F_i(t, iu)|_{t=0} \leq -\frac{u^\top \alpha_i u}{\|u\|^2} \leq -\epsilon$$

uniformly in i and u in the designated set. Also notice the estimate

$$\frac{1}{\|u\|}\mathcal{B}_{ii}\,F_i \leq -\frac{1}{R}|\mathcal{B}_{ii}|\,F_i$$

and the uniform bound on the last summand on the right hand side of (B.6) for $t \leq t_0$ and $\|u\| \geq R$. The claim now follows from Lemma B.3. $\qquad\square$

Below we shall make use of the following is easy to check auxiliary result on Riccati equations:

Lemma B.5. *Let $A > 0, B \in \mathbb{R} \setminus \{0\}$, and $t_0 \geq 0$, and $G_0 < 0$. For $t \geq t_0$, the solution of*

$$\partial_t G(t) = A G(t)^2 + B G(t), \quad G(t_0) = G_0$$

is of the form

$$G(t) = \frac{B G_0 e^{B(t-t_0)}}{(A G_0 + B) - A G_0 e^{B(t-t_0)}}.$$

If $B = 0$, then

$$G(t) = \frac{G_0}{1 - A G_0 (t - t_0)}.$$

From (B.4) we deduce the trivial differential inequality

$$\partial_t F_i \leq \alpha_{i,ii}\,F_i^2 + \frac{\mathcal{B}_{ii}}{\|u\|}\,F_i.$$

By Lemma B.5, the solution of

$$\partial_t h = \alpha_{i,ii}\,h^2 + \frac{\mathcal{B}_{ii}}{\|u\|}\,h$$

with $h(t_0) < 0$ is explicitly given by

$$h(t) = -\frac{e^{\frac{\mathcal{B}_{ii}}{\|u\|}(t-t_0)}}{\|u\|\frac{\alpha_{i,ii}}{\mathcal{B}_{ii}}\left(e^{\frac{\mathcal{B}_{ii}}{\|u\|}(t-t_0)} - 1\right) - \frac{1}{h(t_0)}}, \quad t \geq t_0.$$

Together with Lemmas B.3 and B.4 and we thus obtain that

$$\text{(B.7)} \qquad F_i(t, iu) \leq -\frac{e^{\frac{\mathcal{B}_{ii}}{\|u\|}(t-t_0)}}{\|u\|\frac{\alpha_{i,ii}}{\mathcal{B}_{ii}}\left(e^{\frac{\mathcal{B}_{ii}}{\|u\|}(t-t_0)} - 1\right) + \frac{1}{\rho}}, \quad t \geq t_0$$

for all $\|u\| \geq R$ with $u^\top \alpha_i\, u \geq \epsilon\|u\|^2$. By rescaling we infer

$$f_i(t, iu) = \|u\| F_i(t\|u\|, iu)$$

$$\leq -\frac{\|u\| e^{\mathcal{B}_{ii}\left(t - \frac{t_0}{\|u\|}\right)}}{\|u\|\frac{\alpha_{i,ii}}{\mathcal{B}_{ii}}\left(e^{\mathcal{B}_{ii}\left(t - \frac{t_0}{\|u\|}\right)} - 1\right) + \frac{1}{\rho}}$$

$$= -\frac{1}{\alpha_{i,ii}}\frac{d}{dt}\log\left(\|u\|\frac{\alpha_{i,ii}}{\mathcal{B}_{ii}}\left(e^{\mathcal{B}_{ii}\left(t - \frac{t_0}{\|u\|}\right)} - 1\right) + \frac{1}{\rho}\right), \quad t \geq \frac{t_0}{\|u\|}$$

for all $\|u\| \geq R$ with $u^\top \alpha_i \, u \geq \epsilon \|u\|^2$. Integrating this inequality yields

$$
\int_0^t f_i(s, \mathrm{i}u) \, ds \leq \int_{\frac{t_0}{\|u\|}}^t f_i(s, \mathrm{i}u) \, ds
$$

(B.8)
$$
= -\frac{1}{\alpha_{i,ii}} \left[\log \left(\|u\| \frac{\alpha_{i,ii}}{\mathcal{B}_{ii}} \left(e^{\mathcal{B}_{ii}\left(t - \frac{t_0}{\|u\|}\right)} - 1 \right) + \frac{1}{\rho} \right) - \log \left(\frac{1}{\rho} \right) \right]
$$

$$
= -\frac{1}{\alpha_{i,ii}} \log \left(\rho \|u\| \frac{\alpha_{i,ii}}{\mathcal{B}_{ii}} \left(e^{\mathcal{B}_{ii}\left(t - \frac{t_0}{\|u\|}\right)} - 1 \right) + 1 \right), \quad t \geq \frac{t_0}{\|u\|}
$$

for all $\|u\| \geq R$ with $u^\top \alpha_i \, u \geq \epsilon \|u\|^2$. We arrive at the following key result, which completes the proof of Lemma B.2.

Lemma B.6. *Let $i \in \{1, \ldots, m\}$. For every $\epsilon > 0$ and $t > 0$ there exists some $R > 0$ and $C > 0$ such that*

$$
\left| e^{\phi(t, \mathrm{i}u) + \psi(t, \mathrm{i}u)^\top x} \right| \leq C \left(1 + \|u\| \right)^{-\frac{b_i}{\alpha_{i,ii}}}
$$

for all $u \in \mathbb{R}^d$ with $\|u\| \geq R$ and $u^\top \alpha_i \, u \geq \epsilon \|u\|^2$. Moreover,

$$
\Re \phi(t, \mathrm{i}u) \leq -u_J^\top \left(\int_0^t e^{\mathcal{B}_{JJ}^\top s} \, a \, e^{\mathcal{B}_{JJ} s} \, ds \right) u_J
$$

for all $u \in \mathbb{R}^d$.

Proof. Integration of (4.2) implies

$$
\Re \left(\phi(t, \mathrm{i}u) + \psi(t, \mathrm{i}u)^\top x \right) \leq \Re \phi(t, \mathrm{i}u)
$$

$$
\leq -u_J^\top \left(\int_0^t e^{\mathcal{B}_{JJ}^\top s} \, a \, e^{\mathcal{B}_{JJ} s} \, ds \right) u_J + b_i \int_0^t f_i(s, \mathrm{i}u) \, ds.
$$

Together with (B.8), this proves the lemma. $\qquad\square$

Appendix C. Proof of Theorem 4.3

Fix some $\gamma > 0$. We consider the two-dimensional process $Z = (X, Y)$, where $Y_t = y + \gamma \int_0^t X_s^x ds$, with $y \in \mathbb{R}_+$. It is easy to see that (X, Y) is an affine process with state space $\mathbb{R}_+^2$. In particular, if $y = 0$ we have that $Y_t = \gamma \int_0^t X_s^x$ has an exponentially affine characteristic function of the form

$$
\mathbb{E} \left[e^{\mathrm{i}v Y_t} \mid X_0 = x \right] = e^{\phi(t, \mathrm{i}v) + \psi(t, \mathrm{i}v) x}, \quad v \in \mathbb{R},
$$

where the characteristic exponents ϕ and ψ satisfy the generalized Riccati differential equations

$$
\partial_t \phi = b \psi
$$
$$
\phi(0) = 0
$$
$$
\partial_t \psi = \alpha \psi^2 + \beta \psi + \mathrm{i}\gamma v + \int_0^\infty \left(e^{\psi \xi} - 1 \right) \mu(d\xi)
$$
$$
\psi(0) = 0.
$$

For any $v \neq 0$, we define the scaled functions[16]

$$F(t, iv) = \frac{1}{\sqrt{|v|}} \Re \psi \left(\frac{t}{\sqrt{|v|}}, iv \right)$$

$$G(t, iv) = \frac{1}{\sqrt{|v|}} \Im \psi \left(\frac{t}{\sqrt{|v|}}, iv \right).$$

Then F and G satisfy

$$\partial_t F = \alpha \left(F^2 - G^2 \right) + \frac{\beta}{\sqrt{|v|}} F + \frac{1}{|v|} \int_0^\infty e^{\sqrt{|v|} F \xi} \left(\cos \left(\sqrt{|v|} F \xi \right) - 1 \right) \mu(d\xi)$$

$$F(0) = 0$$

(C.1)

$$\partial_t G = 2\alpha F G + \frac{\beta}{\sqrt{|v|}} G + \gamma \frac{v}{|v|} + \frac{1}{|v|} \int_0^\infty e^{\sqrt{|v|} F \xi} \sin \left(\sqrt{|v|} F \xi \right) \mu(d\xi)$$

$$G(0) = 0$$

We now prove a first intermediary result, which is analogous to Lemma B.4:

Lemma C.1. *There exists some $t_0 > 0$, $\rho > 0$ and $R > 0$ such that*

$$F(t_0, iv) \leq -\rho$$

for all v with $|v| \geq R$.

Proof. For $v \to \pm\infty$, the solutions $F(t, iv)$ and $G(t, iv)$ of (C.1) converge locally uniformly in t to the solutions $F_\infty(t)$ and $G_\infty(t)$ of the system

$$\partial_t F_\infty = \alpha \left(F_\infty^2 - G_\infty^2 \right)$$

$$F_\infty(0) = 0$$

$$\partial_t G_\infty = 2\alpha F_\infty G_\infty \pm 1$$

$$G_\infty(0) = 0.$$

Since $\partial_t G_\infty(t)|_{t=0} = \pm 1$ it follows that there exists some $t_1 > 0$ such that $G_\infty(t) \neq 0$ for all $t \in (0, t_1)$. This again implies that $F_\infty(t_0) < 0$ for some $t_0 \in (0, t_1)$, and the lemma follows. $\square$

From (C.1) we deduce the trivial differential inequality

$$\partial F \leq \alpha F^2 + \frac{\beta}{\sqrt{|v|}} F.$$

Arguing as for the derivation of (B.7), we then obtain together with Lemmas B.3 and C.1 that

$$F(t, iv) \leq -\frac{e^{\frac{\beta}{\sqrt{|v|}}(t-t_0)}}{\sqrt{|v|} \frac{\alpha}{\beta} \left(e^{\frac{\beta}{\sqrt{|v|}}(t-t_0)} - 1 \right) + \frac{1}{\rho}}, \quad t \geq t_0$$

---

[16]Note that here we have to scale by $\sqrt{|v|}$, which is in contrast to the proof of Theorem 4.1, see (B.3).

for all $v \in \mathbb{R}$ with $|v| \geq R$. By rescaling and integrating we infer, arguing as for the derivation of (B.8), that

$$\int_0^t \Re\psi(s, iv)\, ds \leq \int_{\frac{t_0}{\sqrt{|v|}}}^t \Re\psi(s, iv)\, ds$$

$$= -\frac{1}{\alpha}\left[\log\left(\sqrt{|v|}\frac{\alpha}{\beta}\left(e^{\beta\left(t - \frac{t_0}{\sqrt{|v|}}\right)} - 1\right) + \frac{1}{\rho}\right) - \log\left(\frac{1}{\rho}\right)\right]$$

$$= -\frac{1}{\alpha}\log\left(\rho\sqrt{|v|}\frac{\alpha}{\beta}\left(e^{\beta\left(t - \frac{t_0}{\sqrt{|v|}}\right)} - 1\right) + 1\right), \quad t \geq \frac{t_0}{\sqrt{|v|}}$$

for all $v \in \mathbb{R}$ with $|v| \geq R$. Similarly as in Lemma B.6 we now infer that

$$\left|e^{\phi(t, iv) + \psi(t, iv)x}\right| \leq C\left(1 + |v|\right)^{-\frac{b}{2\alpha}}.$$

Combining this with Lemma B.1 completes the proof of Theorem 4.3.

Appendix D. Figures and Tables

	Heston Model			BAJD	
KS	Order 2	Order 4	KS	Order 2	Order 4
$\kappa\theta_V$	0.0199	0.6654	$\kappa\theta_V$	0.0038	0.2360
κ_V	0.0000	0.0693	κ_V	0.0396	0.7658
σ_V	0.0001	0.2574	σ_V	0.0000	0.6754
$\kappa\theta_X$	0.0018	0.0348	l	0.0000	0.0571
ρ	0.0003	0.3291	ν	0.0000	0.0049

TABLE 2. **Kolmogorov-Smirnov test statistics:** The table displays p-values for a two-sided Kolmogorov-Smirnov test applied to posterior density p_{VX} to $p_{VX}^{(2)}$ and $p_{VX}^{(4)}$ using prior (7.21) for the Heston model in the left panel. The right panel displays p-values for the test applied to p_Y to $p_Y^{(2)}$ and $p_Y^{(4)}$, the BAJD model. The prior distribution for this model is defined in eq. (7.20). The true posterior is defined in eq. (7.18) and the approximate posterior densities are defined in eq. (7.19).

	MLE	BG(4)	QML	G(4)	CF(2)
#Success	688	841	982	949	981

TABLE 3. **Heston Estimation Success:** The table reports the number of estimation successes on 1,000 datasets generated as exact draws from the Heston model using the technology from Broadie and Kaya (2006). Estimation success is defined by the optimizer meeting the termination criterion, which is a function of the norm of the gradient of the log likelihood function. The density approximations used are BG(4), a fourth order expansion using a Bilateral Gamma weight for the log stock variable and a Gamma weight for the variance variable, G(4), a fourth order expansion using a Gaussian weight for the log stock variable and a Gamma weight for the variance variable, QML denotes a Gaussian approximation using the true conditional moments up to order 2, and CF(2) denotes the second-order likelihood expansions from Aït-Sahalia (2008). The optimizer used in the likelihood search is donlp2.

(A) Bias

	ϱ_{VX}^{TRUE}	MLE		$BG(4)$		QML		$G(4)$		$CF(2)$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
κ	1	0.2255	0.4253	0.2313	0.4168	0.2324	0.4214	0.2327	0.4187	0.1905	0.5553
$\kappa\theta_V$	0.04	0.0093	0.0166	0.0094	0.0164	0.0093	0.0167	0.0095	0.0164	0.0063	0.0179
σ	0.2	0.0009	0.0049	0.0007	0.0047	0.0008	0.0047	0.0009	0.0047	0.0044	0.0117
$\kappa\theta_X$	0.03	−0.0151	0.0473	−0.0145	0.0469	−0.0138	0.0483	−0.0150	0.0472	−0.0050	0.0624
ρ	−0.8	−0.0016	0.0147	−0.0001	0.0138	−0.0015	0.0140	−0.0016	0.0138	0.0003	0.0270

(B) Estimation Noise

	ϱ_{VX}^{TRUE}	$\widehat{\varrho}_{VX}^{\star MLE} - \varrho_{VX}^{TRUE}$		$\widehat{\varrho}_{VX}^{\star BG(4)} - \widehat{\varrho}_{VX}^{\star MLE}$		$\widehat{\varrho}_{VX}^{\star QML} - \widehat{\varrho}_{VX}^{\star MLE}$		$\widehat{\varrho}_{VX}^{\star G(4)} - \widehat{\varrho}_{VX}^{\star MLE}$		$\widehat{\varrho}_{VX}^{\star CF(2)} - \widehat{\varrho}_{VX}^{\star MLE}$	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
κ	1	0.2255	0.3609	0.0057	0.1235	0.0069	0.1313	0.0072	0.1179	−0.0350	0.4336
$\kappa\theta_V$	0.04	0.0093	0.0137	0.0001	0.0024	−0.0001	0.0043	0.0002	0.0025	−0.0030	0.0092
σ	0.2	0.0009	0.0048	−0.0002	0.0014	−0.0001	0.0015	0.0000	0.0014	0.0035	0.0100
$\kappa\theta_X$	0.03	−0.0151	0.0448	0.0006	0.0121	0.0013	0.0171	0.0001	0.0129	0.0101	0.0390
ρ	−0.8	−0.0016	0.0146	0.0015	0.0049	0.0001	0.0050	−0.0001	0.0049	0.0018	0.0255

TABLE 4. **Heston Asymptotic Assessment:** Panel (a) displays bias and RMSE of the MLE estimator and the approximated MLE estimators. Panel (b) displays mean and standard deviation of ML estimation bias as well as the first two moments of the difference between the MLE estimator and the approximated MLE estimators. Computed over a sample of 1,000 datasets, all of which generated as exact draws from the Heston model using the technology from Broadie and Kaya (2006). For a given dataset only parameter estimates were taken into consideration where all five estimators converged. Out of 1,000 this left 578 samples. The number of estimation successes is reported in Table 3 above. Approximate estimators are obtained through BG(4), a fourth order expansion using a Bilateral Gamma weight for the log stock variable and a Gamma weight for the variance variable, G(4), a fourth order expansion using a Gaussian weight for the log stock variable and a Gamma weight for the variance variable, QML denotes a Gaussian approximation using the true conditional moments up to order 2, and CF(2) denotes the second-order likelihood expansions from Aït-Sahalia (2008).
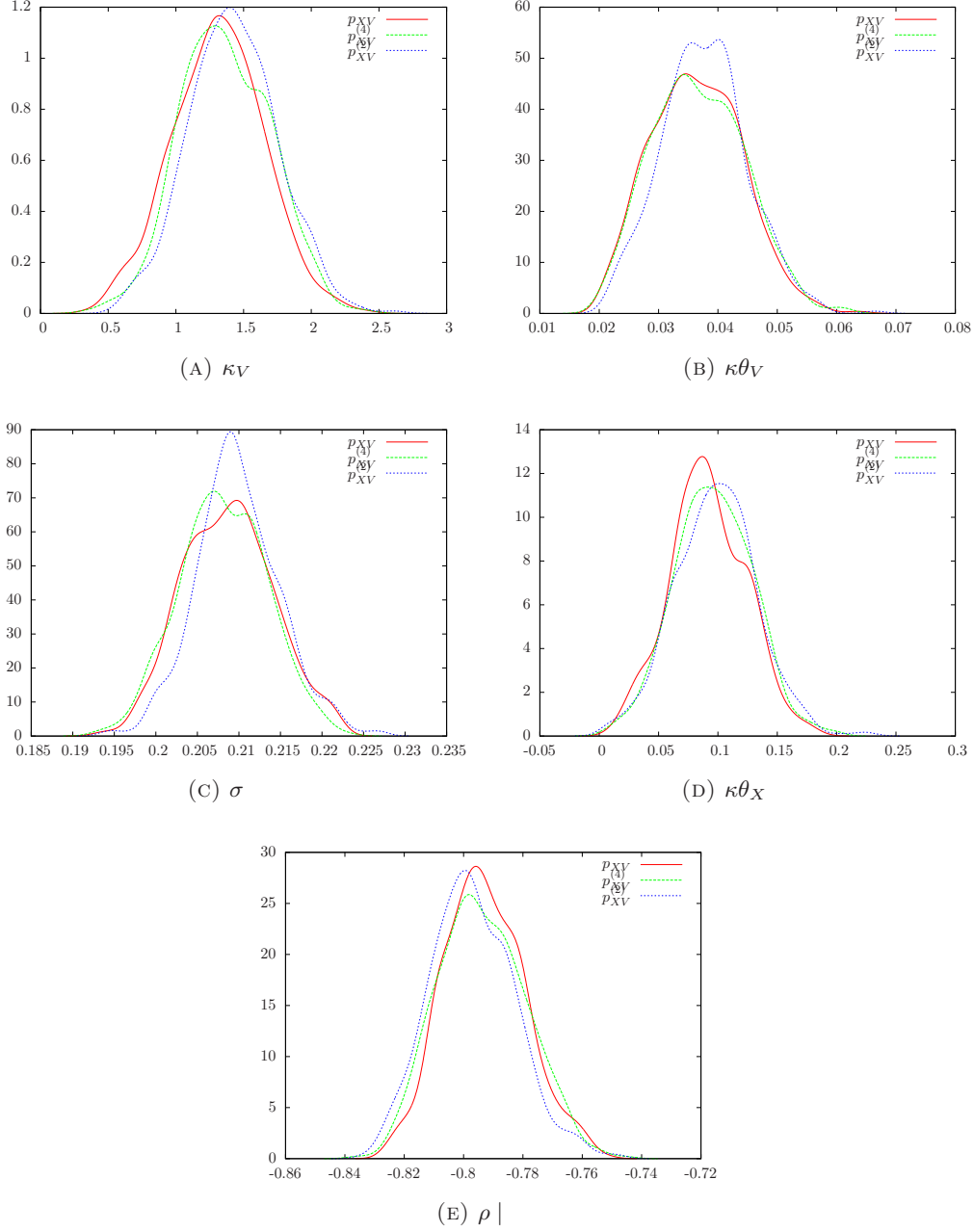
FIGURE 5. **Posterior Densities for Heston's model:** The figure displays the marginal posterior distributions of the parameters of Heston's model conditional on the data. Bayesian estimation is performed using prior specification (7.21) with the true transition density g_{VX} obtained through Fourier inversion, closed-form density up to second $(g_{VX}^{(2)})$, and fourth order $(g_{VX}^{(4)})$.

References

AÏT-SAHALIA, Y. (2002): "Maximum likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.

——— (2007): *Estimating Continuous-Time Models Using Discretely Sampled Data*, Cambridge University Press: in Richard Blundell, Torsten Persson, Whitney K. Newey: *Advances in Economics and Econometrics, Theory and Applications*, chapter 9.

——— (2008): "Closed-Form Likelihood Expansions for Multivariate Diffusions," *Annals of Statistics*, 36, 906–937.

AÏT-SAHALIA, Y. AND L. P. HANSEN, eds. (2009): *Handbook of Financial Econometrics*, Elsevier.

AÏT-SAHALIA, Y. AND R. KIMMEL (2007): "Maximum likelihood estimation of stochastic volatility models," *Journal of Financial Economics*, 83, 413–452.

AÏT-SAHALIA, Y. AND J. YU (2005): "Saddlepoint Approximations for Continuous-Time Markov Processes," *Journal of Econometrics*, forthcoming.

ANDERSEN, L., J. SIDENIUS, AND S. BASU (2003): "All your hedges in one basket," *Risk*.

BATES, D. (2006): "Maximum Likelihood Estimation of Latent Affine Processes," *Review of Financial Studies*, 19, 909–965.

BERNARD, P. (1995): *Lecture Notes in Physics*, Springer, vol. 451/1995, chap. Some Remarks Concerning Convergence of Orthogonal Polynomial Expansions, 327–334.

BIBBY, B., M. JACOBSEN, AND M. SØRENSEN (2004): "Estimating functions for discretely sampled diffusion-type models," Working paper, CAF Centre for Analytical Finance, University of Aarhus.

BROADIE, M. AND Ö. KAYA (2006): "Exact Simulation of Stochastic Volatility and other Affine Jump Diffusion Processes," *Operations Research*, 54, 217–231.

BRU, M.-F. (1991): "Wishart Processes," *Journal of Theoretical Probability*, 4, 725–751.

BURASCHI, A., A. CIESLAK, AND F. TROJANI (2008): "Correlation Risk and the Term Structure of Interest Rates," Working paper, Imperial College and University of St. Gallen.

CARR, P. AND D. MADAN (1999): "Option valuation using the fast Fourier transform." *Journal of Computational Finance*, 2, 61–73.

CHEN, H. AND S. JOSLIN (2011): "Generalized Transform Analysis of Affine Processes and Applications in Finance," Working paper, Sloan School of Management and Marshall School of Business.

COLLIN-DUFRESNE, P., R. S. GOLDSTEIN, AND C. S. JONES (2008): "Identification of Maximal Affine Term Structure Models," *Journal of Finance*, 63, 743–795.

COX, J., J. INGERSOLL, AND S. ROSS (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

CUCHIERO, C., D. FILIPOVIĆ, E. MAYERHOFER, AND J. TEICHMANN (2010a): "Affine Processes on Positive Semidefinite Matrices," *The Annals of Applied Probability*, 21, 397–463.

CUCHIERO, C., J. TEICHMANN, AND M. KELLER-RESSEL (2010b): "Polynomial Processes and their application to mathematical Finance," *Finance & Stochastics*, forthcoming.

DA FONSECA, J., M. GRASSELLI, AND C. TEBALDI (2008): "A multifactor volatility Heston model," *Quantitative Finance*, 8, 591 – 604.

DAI, Q. AND K. J. SINGLETON (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance*, 55, 1943–1978.

DI PIETRO, M. (2001): "Bayesian Inference for Discretely Sampled Diffusion Processes with Financial Applications," Ph.D. thesis, Carnegie Mellon University.

DUFFIE, D., D. FILIPOVIĆ, AND W. SCHACHERMAYER (2003): "Affine Processes and Applications in Finance," *Annals of Applied Probability*, 13, 984–1053.

DUFFIE, D. AND N. GARLEANU (2001): "Risk and Valuation of Collateralized Debt Obligations," *Financial Analysts Journal*, 57, 41–59.

DUFFIE, D. AND R. KAN (1996): "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, 6, 379–406.

DUFFIE, D., J. PAN, AND K. SINGLETON (2000): "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

ECKNER, A. (2009): "Computational Techniques for basic Affine Models of Portfolio Credit Risk," *Journal of Computational Finance*, 100, 1 – 35.

ELERIAN, O., S. CHIB, AND N. SHEPHARD (2001): "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69, 959–993.

ERAKER, B. (2001): "MCMC Analysis of Diffusion Models with Application to Finance," *Journal of Business & Economic Statistics*, 19, 177–191.

——— (2004): "Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices," *Journal of Finance*, 59, 1367–1404.

ERAKER, B., M. JOHANNES, AND N. POLSON (2003): "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 58, 1269–1300.

FELDHÜTTER, P. (2008): "An Empirical Investigation of an Intensity-Based Model for Pricing CDO Tranches," Working paper, Copenhagen Business School.

FILIPOVIĆ, D. (2009): *Term-Structure Models: A Graduate Course*, Springer, Berlin.

FILIPOVIĆ, D. AND E. MAYERHOFER (2009): "Affine diffusion processes: theory and applications," .

FORMAN, J. L. AND M. SØRENSEN (2008): "The Pearson Diffusions: A Class of Statistically Tractable Diffusion Processes," *Scandinavian Journal of Statistics*, 35, 438–465.

GALLANT, A. R. AND G. TAUCHEN (2009): *Simulated Score Methods and Indirect Inference for Continuous-time Models*, in Aït-Sahalia and Hansen (2009).

HESTON, S. (1993): "A closed-form solution for options with stochastic volatility with applications to bond and currency options," *Review of Financial Studies*, 6, 327–343.

HORN, R. A. AND C. R. JOHNSON (1990): *Matrix analysis*, Cambridge: Cambridge University Press, corrected reprint of the 1985 original.

HURN, A., J. JEISMAN, AND K. LINDSAY (2007): "Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations," *Journal of Financial Econometrics*, 5, 390–455.

——— (2008): "Horses for courses: Polynomial-based approximations of transitional density," Working paper, Queensland University of Technology, University of Glasgow.

JOHNSON, N. L., S. KOTZ, AND N. BALAKRISHNAN (1995): *Continuous univariate distributions. Vol. 2*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, New York: John Wiley & Sons Inc., second ed., a Wiley-Interscience Publication.

JONES, C. S. (1998): "Bayesian Estimation of Continuous-Time Finance Models," Working paper, University of Rochester.

KRISTENSEN, D. AND A. MELE (2011): "Adding and subtracting Black-Scholes: A new approach to approximating derivativeprices in continuous-time models," *Journal of Financial Economics*, forthcoming.

KÜCHLER, U. AND S. TAPPE (2008a): "Bilateral Gamma distributions and processes in financial mathematics," *Stochastic Processes and Their Applications*, 118, 261 – 283.

——— (2008b): "On the shapes of bilateral Gamma densities," *Statistics and Probability Letters*, 78, 2478 – 2484.

LAMOUREUX, C. G. AND A. PASEKA (2005): "Information in Options and Underlying Asset Dynamics," *working paper, University of Arizona*.

LANDO, D. (1998): "On Cox Processes and Credit Risky Securities," *Review of Derivatives Research*, 2, 99–120.

LE, A., K. SINGLETON, AND Q. DAI (2010): "Discrete-Time Affine$^\mathbb{Q}$ Term Structure Models with Generalized Market Prices of Risk," *Review of Financial Studies*, 23, 2184–2227.

LEIPPOLD, M. AND F. TROJANI (2008): "Asset Pricing with Matrix Jump Diffusions," Tech. rep., Swiss Finance Institute University of Zürich and University of Lugano.

MORTENSEN, A. (2006): "Semi-Analytical Valuation of Basket Credit Derivatives in Intensity-Based Models," *Journal of Derivatives*, 13, 8–28.

ROBERT, C. AND G. CASELLA (2004): *Monte Carlo Statistical Methods*, New York: Springer.

ROBERT, C. P. (1994): *The Bayesian Choice*, New York: Springer.

ROBERTS, G. O. AND O. STRAMER (2001): "On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm," *Biometrika*, 88, 603–621.

ROGERS, L. (1985): "Smooth Transition Densities for One-Dimensional Diffusions," *Bull. London Math. Soc.*, 17, 157–161.

SATO, K.-I. (1999): *Lévy processes and infinitely divisible distributions*, vol. 68 of *Cambridge Studies in Advanced Mathematics*, Cambridge: Cambridge University Press, translated from the 1990 Japanese original, Revised by the author.

SCHNEIDER, P., L. SÖGNER, AND T. VEŽA (2010): "The Economic Role of Jumps and Recovery Rates in the Market for Corporate Default Risk," *Journal of Financial and Quantitative Analysis*, 45, 1517–1547.

SCHOUTENS, W. (2000): *Stochastic Processes and Orthogonal Polynomials*, vol. 146 of *Lecture Notes in Statistics*, New York: Springer.

SINGLETON, K. (2001): "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111–141.

SØRENSEN, H. (2004): "Parametric Inference for Diffusion Processes Observed at Discrete Points: a Survey," *International Statistical Review*, 72, 337–354.

STRAMER, O., M. BOGNAR, AND P. SCHNEIDER (2009): "Bayesian Inference for Discretely Sampled Markov Processes with closed–form Likelihood Expansions," *Journal of Financial Econometrics*, forthcoming.

VASICEK, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

VOLKMANN, P. (1972): "Gewöhnliche Differentialungleichungen mit quasimonoton wachsenden Funktionen in topologischen Vektorräumen," *Math. Z.*, 127, 157–164.

YU, J. (2007): "Closed-Form Likelihood Estimation of Jump-Diffusions with an Application to the Realignment Risk of the Chinese Yuan," *Journal of Econometrics*, 141, 1245–1280.